Identification and Functional Analysis of Single Nucleotide Polymorphisms that affect Human Cancer

Lukasz Filip Grochola, MD

A thesis submitted in fulfillment of the requirements for the degree of

Doctor of Philosophy

Trinity Term 2011

Word count: 49,500

Hertford College
Ludwig Institute for Cancer Research
Nuffield Department of Clinical Medicine
University of Oxford

Abstract

Identification and Functional Analysis of Single Nucleotide Polymorphisms that affect Human Cancer

Lukasz Filip Grochola, Hertford College. Doctor of Philosophy, Trinity Term 2011.

Aims: The p53 regulatory network is crucial in directing the suppression of cancer formation and mediating the response to commonly used cancer therapies. Functional genetic variants in the genes comprising this network could help identify individuals at greater risk for cancer and patients with poorer responses to therapies, but few such variants have been identified as yet. Methods: We first develop and apply three different screens that utilize known characteristics of functional single nucleotide polymorphisms (SNPs) in the p53 network to search for variants that associate with allelic differences in (i) recent natural selection, (ii) chemosensitivity profiles, and (iii) the gender- and age-dependent incidence of soft-tissue sarcoma. Secondly, we study and explore the functional mechanisms associated with the identified variants. Results: We identify SNPs in the PPP2R5E, CD44, YWHAQ and ESR1 genes that associate with allelic differences in the age of tumour diagnosis (up to 32.5 years, $p=0.031$), cancer risk (up to 8.1 odds ratio, $p=0.004$) and overall survival (up to 2.85 relative risk, $p=0.011$) in sarcomas, ovarian and pancreatic cancers, and exhibit allelic differences in the cellular responses to cytotoxic chemotherapeutic agents (up to 5.4-fold, $p=5.6\times10^{-47}$). Lastly, we identify candidate causal SNPs in those genes and describe the regulatory mechanisms by which they might affect human cancer. Conclusions: Together, our work suggests that the inherited genetics of the p53 pathway have a great potential to further define populations in their abilities to react to stress, suppress tumor formation and respond to therapies.

Abbreviations

- Amp: Amplitude
- APS: Ammonium persulfate
- AR: Androgen receptor

- bHLH: basic-Helix-Loop-Helix
- Bp: Base pair
- BSA: Bovine serum albumin

- cDNA: Complementary DNA
- CEU: Individuals from the Centre d'Etude du Polymorphisme Humain collected in Utah, USA, with ancestry from northern and western Europe
- ChIP: Chromatin immunoprecipitation
- CN: Copy number
- CPM: Counts per minute
- CSC: Cancer stem cells

- ΔCt: delta Cycle-threshold
- dATP: Deoxyadenosine triphosphate
- ddH2O: Double-distilled water
- dGTP: Deoxyguanosine triphosphate

- DMEM: Dulbecco's Modified Eagle's Medium
- DMSO: Dimethyl sulfoxide
- DNA: Deoxyribonucleic acid
- dNTP: Deoxyribonucleotide
- DTP: Developmental Therapeutics Program
- DTT: Dithiothreitol
- dTTP Deoxythymidine triphosphate

- ECL: Enhanced Chemical Luminescence
- EDTA: Ethylenediaminetetraacetic acid
- EMSA: Electromobility shift assays
- ER: Estrogen receptor
- ERα: Estrogen receptor alpha
- ESE: Exonic splicing enhancers
- ESR1: Estrogen receptor alpha
- ESS: Exonic splicing silencers

- FBS: Foetal Bovine Serum

- GAPDH: Glyceraldehyde 3-phosphate dehydrogenase
- GWAS: Genome-wide association study
- Gy: Gray

- HCl: Hydrochloric acid
- Hr: Hour

- Ig: Immunoglobulin
- IGFR1: Insulin-like growth factor receptor 1
- iHS: Integrated haplotype score
- IP: Immunoprecipitate

- Kb: Kilo base pairs
- KCl: Potassium chloride

- L: Liter
- LD: Linkage disequilibrium
- LFS: Li Fraumeni Syndrome
- LIF: Leukemia inhibitory factor

- MAF: Minor allele frequency
- MDM2: Murine double minute 2
- MDM4: Murine double minute 4
- MgCl2: Magnesium chloride
- Mo: Months

- NaCl: Sodium chloride

- NaH2PO4: Sodium dihydrogen phosphate
- NCI: National Cancer Institute
- NOS: Non otherwise specified

- P/S: Penicillin/Streptomycin
- PAGE: Polyacrylamide gel electrophoresis
- PAI-1: Plasminogen activator inhibitor 1
- PBS: Phosphate Bufferd Saline (PBS)
- PCR: Polymerase Chain Reaction
- PDAC: Pancreatic ductal adenocarcinoma
- PI: Protease inhibitors
- PI3K: Phosphatidylinositol 3-kinase
- Pol II: RNA Polymerase II
- PP2A: Protein phosphatase 2A
- PPP2R5E: Protein phosphatase 2, regulatory subunit B′, ε-isoform

- R0: Complete resection with no residual tumor
- R1: Microscopic residual disease
- R2: Macroscopic residual disease
- RE: Response elements
- RNA: Ribonucleic acid
- Rpm: Revolutions per minute
- RPMI: Roswell Park Memorial Institute medium

- RR: Relative risk
- RT-PCR: Real-time reverse-transcription PCR

- SDS: Sodium dodecyl sulphate
- SDS-PAGE: Sodium dodecyl sulphate polyacrylamide gel electrophoresis
- Sec: Second
- SNP: Single nucleotide polymorphism
- STS: Soft-tissue sarcoma

- TBE: Tris/Borate/EDTA
- TEMED: Tetramethylethylenediamine (TEMED)
- TF: Transcription factor
- TK: Thymidine kinase
- TSS: Transcriptional start site

- UVB: Ultraviolet B (medium wave)

- v/v: Volume per volume
- vs: Versus

- w/v: Weight per volume

- Y/Yr: Year

I. General Introduction

1.1. The p53 network of genes

1.1.1. p53 in the cell

Since its first description in 1979, the p53 protein and its cellular network of genes has constituted one of the most extensively studied areas in cancer biology (Lane and Crawford 1979; Linzer and Levine 1979; Lane and Levine 2010). In more than 3 decades of intense research, compelling evidence has established the role of the p53 pathway as a major cellular tumor suppressor network (Lane and Levine 2010). The p53 protein has been demonstrated to suppress cancer primarily through the induction of cell-cycle-arrest, apoptosis or senescence in response to a wide range of different cellular stresses in a stimulus- and cell-type dependent manner (Riley, Sontag et al. 2008; Vousden and Prives 2009; Lane and Levine 2010; Lu 2010). The stress signals include genotoxic damage, hypoxia, loss of normal cell contact, nutrient deprivation, mitotic spindle damage, heat or cold shock, telomere shortening, unfolded proteins and oncogene activation (Riley, Sontag et al. 2008; Vousden and Prives 2009). Upon cellular stress, the p53 protein undergoes a plethora of post-translational modifications, which can affect the stability and activity of the protein (Levine, Hu et al. 2006; Vousden and Prives 2009). Specifically, p53 is phosphorylated on a large number of serine and threonine residues by many different protein kinases, acetylated by histone acetyl-transferases, methylated by methylases, ubiquitinated, summolated and neddylated (Brooks and Gu 2003; Xu 2003; Xirodimas, Saville et al. 2004; Levine, Hu et al. 2006; Li, Piluso et al. 2006). It has been shown that these processes can mediate the response from cellular stress signaling to the

p53 protein (Brooks and Gu 2003; Xu 2003; Xirodimas, Saville et al. 2004; Levine, Hu et al. 2006). For example, in response to DNA damage from ionizing or UV radiation, p53 is phosphorylated at the amino terminus at specific amino acids by various kinases, including ATM, ATR, DNA-PK, CHK1, and CHK2 (Levine, Hu et al. 2006).

Following activation, p53 is stabilized and activated and binds to DNA as a tetramer in a sequence-specific manner, which leads to the transcriptional regulation of the downstream genes of the p53 network (Beckerman and Prives 2010). The exact downstream target gene is dictated by cell type, environmental milieu, the nature as well as severity of the stress (Beckerman and Prives 2010; Lu 2010). A set of genes is involved in cell cycle arrest, most prominently p21 Waf1/Cip1, 14-3-3 sigma and GADD-45 (Levine, Hu et al. 2006; Vousden and Prives 2009). For example, activation of p53 upon mild DNA-damage has been shown to induce a G1 arrest through the transactivation of p21Waf1/Cip1 (CDKN1A), a cyclin-dependent kinase inhibitor, which delays the progression into the next cell cycle phase until the DNA damage has been repaired (Levine, Hu et al. 2006; Vousden and Prives 2009). Other p53-regulated genes function as inducers of the intrinsic apoptotic pathway, such as bax, noxa and puma, which promote cytochrome c release in different cell types (Levine, Hu et al. 2006; Vousden and Prives 2009). In the extrinsic apoptotic pathway, p53 induces, among other targets, the transcription of Fas, killer/DR5 and the trail receptor (Levine, Hu et al. 2006; Vousden and Prives 2009). These proteins, together with PIDD, activate caspase 8 and Bid to release cytochrome c, which acts with the p53-regulated gene APAF-1 to activate caspase 9 and then 3, thus leading to apoptosis and ensuring an intrinsic fail-safe mechanism to resist cellular transformation after potent stress signals (Levine, Hu et al.

2006; Vousden and Prives 2009). Besides the two extensively studied p53 stress responses cell cycle arrest and apoptosis, the induction of senescence has also been implicated in the tumor-suppressive role of p53 (Levine, Hu et al. 2006; Vousden and Prives 2009; Zilfou and Lowe 2009). Cellular senescence is a permanent and irreversible form of cell-cycle arrest and can be activated by p53 upon induction by various stressors, including dysfunctional telomeres, nontelomeric DNA damage, excessive mitogenic signaling, and perturbations in chromatin organization (Zilfou and Lowe 2009). As with other p53 responses, senescence results from changes in the expression of a number of proteins such as the plasminogen activator inhibitor 1 (PAI-1) or, in a context-dependent fashion, by p21 (Brown, Wei et al. 1997; Kortlever, Higgins et al. 2006; Leal, Fominaya et al. 2008; Zilfou and Lowe 2009). In addition, other regulatory functions of p53 in tumor suppression have been discovered, which do not depend on its ability to transactivate gene expression. Specifically, p53 can also directly regulate proteins from the Bcl-2 family, (Moll, Wolff et al. 2005), microRNA processing (Suzuki, Yamagata et al. 2009) and lincRNA expression (Huarte, Guttman et al. 2010).

The complex regulation of p53 enables it to integrate a variety of stress signals into specific cellular responses in a stimulus and cell-type dependant manner. An interesting component of this regulation is a multitude of feedback loops, which positively or negatively affect the activity of p53 (Lu 2010). Hereby, the MDM2-p53 feedback loop is thought to be the key regulatory component of the p53 stress response (Lu 2010). It controls the level of expression and the activity of p53 by different mechanisms (Lu 2010; Manfredi 2010). Murine double minute 2 (MDM2) is an E3 ubiquitin ligase, which can bind the p53 protein and target it for proteosome-mediated

degradation through ubiquitination (Lu 2010; Manfredi 2010). Furthermore, MDM2 directly binds to the N-terminus of p53, thereby masking its transactivation domain (Manfredi 2010). In addition, MDM2 has been demonstrated to shuttle p53 out of the nucleus, where it can no longer function as a transcriptional activator (Wade, Wang et al. 2010). Interestingly, MDM2 itself is a transcriptional target of p53, thus resulting in a negative feedback loop between p53 and MDM2 (Lu 2010; Manfredi 2010). A variety of stress signals can lead to the disruption of the interaction between MDM2 and p53, resulting in the stabilization and activation of p53 and its downstream cellular responses (Riley, Sontag et al. 2008). For example, after treatment with ionizing radiation, both MDM2 and p53 are post-translationally modified through phosphorylation by the ATM kinase, which disrupts the interaction with p53 and, furthermore, inhibits the E3 ligase activity of MDM2, resulting in the stabilization of p53, an increase in p53 protein levels and activation of the p53 stress response (Banin, Moyal et al. 1998; Levine, Hu et al. 2006; Cheng, Chen et al. 2009). Another central node involved in the regulation of the activity of p53 is murine double minute 4 (MDM4), a structural homologue of MDM2 (Wade, Wang et al. 2010). In close similarity to MDM2, the MDM4 protein can bind to the N-terminal transactivation domain of p53 and suppress its transactivating function and thus attenuate the p53 stress response (Wade, Wang et al. 2010). In addition, it has been shown that MDM4 can modulate the translocation of p53 by MDM2 from the nucleus to the cytoplasm (Wade, Wang et al. 2010). In contrast to MDM2, the C-terminal RING domain of MDM4 lacks the ubiquitinase function of MDM2 and therefore cannot by itself target p53 for proteosomal degradation (Linares, Hengstermann et al. 2003). However, it can stimulate MDM2-mediated ubiquitination and degradation of p53

(Linares, Hengstermann et al. 2003) and thus contribute to the attenuation of the tumor suppressive activity of p53. The crucial role of the MDM proteins in the regulation of p53 is further highlighted by the fact that germline inactivation of either MDM2 or MDM4 in mice leads to embryonic lethal phenotypes through an increased activity of p53 during the early stages of development (Jones, Roe et al. 1995; Montes de Oca Luna, Wagner et al. 1995; Parant, Chavez-Reyes et al. 2001; Migliorini, Lazzerini Denchi et al. 2002; Marine, Francoz et al. 2006). Importantly, those phenotypes are completely rescued by concomitant inactivation of the p53 gene, further supporting the notion that MDM2 and MDM4 function mainly through the regulation of the p53 protein (Jones, Roe et al. 1995; Montes de Oca Luna, Wagner et al. 1995; Parant, Chavez-Reyes et al. 2001; Migliorini, Lazzerini Denchi et al. 2002; Marine, Francoz et al. 2006).

1.1.2. p53 in the organism: cancer

The impairment of this tightly regulated network of genes centered on p53 is a hallmark of tumorigenesis (Lane and Levine 2010; Lu 2010). In fact, it has been estimated that 50% of human cancers carry somatic mutations of the p53 gene, which inactivate the normal function of p53 (Levine, Hu et al. 2006; Lane and Levine 2010). In addition, human tumors that retain a p53 wild-type gene frequently show a partial abrogation of downstream effectors (e.g. Bax, Bak, and Apaf-1) or upstream (e.g. ATM, CHK2, MDM2, and p19ARF) regulators of this network (Momand, Jung et al. 1998; Momand, Wu et al. 2000; Johnstone, Ruefli et al. 2002; Danovi, Meulmeester et al. 2004; Lane and Levine 2010; Wade, Wang et al. 2010). For example, the expression of MDM2 and MDM4, the above-described core regulators of p53, is frequently up-regulated in

human cancers in tumors that retain a wild-type 53 gene (Momand, Jung et al. 1998; Momand, Wu et al. 2000; Danovi, Meulmeester et al. 2004). Specifically, MDM2 has been found overexpressed in a variety of human tumors and tumor cell lines as a consequence of gene amplification, which has been noted in 7% of tumors, increased transcript levels, or enhanced translation (Momand, Jung et al. 1998; Momand, Wu et al. 2000). Importantly, the gene amplification of MDM2 occurs predominantly in tumors with a wild-type p53 (Momand, Jung et al. 1998; Momand, Wu et al. 2000). Accordingly, MDM4 has been found overexpressed in a significant percentage of various human tumors and amplified in 5% of primary breast tumors, all of which retained wild-type p53 (Danovi, Meulmeester et al. 2004).

Genes that constitute the p53 network do not only affect the susceptibility to cancer, but are also crucial in the mediation of the cellular response to treatment with DNA-damaging therapeutic agents. Those agents, such as radio- and chemotherapy, lead to the activation of the upstream effectors of p53 (e.g. ATM, ATR, CHK1, CHK2) (Zhou and Elledge 2000; Bolderson, Richard et al. 2009) that activate p53 and can ultimately lead to cancer cell-clearance via activation of apoptotic or senescence pathways (Lowe and Lin 2000; Johnstone, Ruefli et al. 2002). Indeed, mutations in the p53 gene can produce multidrug resistance in cells and in mice, and reintroduction of wild-type p53 can re-constitute chemosensitivity (Wallace-Brodeur and Lowe 1999; Johnstone, Ruefli et al. 2002). By the same token, mutations or altered expression of other p53 network genes, such as APAF-1 or Bcl-2, can significantly alter drug sensitivity in experimental models and are associated with multidrug resistance in human cancers (Reed 1999;

Schmitt, Rosenthal et al. 2000; Zhang, Yu et al. 2000; Wei, Zong et al. 2001; Johnstone, Ruefli et al. 2002).

1.1.3. p53 in the organism: reproduction

Besides its crucial role in maintaining genomic stability and preventing tumor formation, a growing body of evidence suggests that p53 also regulates reproductive processes and fecundity (Hu, Feng et al. 2008). P53 is conserved from invertebrates to mammals (Belyi, Ak et al. 2010) and homologs of p53 have been described in many different organisms, such as sea anemone, clams, Caenorhabditis elegans, Drosophila, frogs, and zebra fish (Soussi, Caron de Fromentel et al. 1987; Cheng, Ford et al. 1997; Jin, Martinek et al. 2000; Derry, Putzke et al. 2001; Kelley, Winge et al. 2001; Pankow and Bamberger 2007; Hu 2009). Interestingly, one of the primary functions of the p53 ancestral gene in worms and flies, which are short-lived, cancer-free organisms, seems to be to ensure germ-line genomic integrity and the fidelity of the developmental process by regulating germ cell replication in order to eliminate defective offspring from the population (Ollmann, Young et al. 2000; Derry, Putzke et al. 2001; Hu, Feng et al. 2008) These primordial p53 activities in the surveillance of germ cell replication and fecundity are also retained in higher organisms. For example, in mice and rats, p53 levels are very high during spermatogenesis and mice with reduced levels of p53 show germ-cell degeneration during the meiotic prophase (Rotter, Schwartz et al. 1993). Furthermore, p53 has also been shown to regulate reproduction in mice, whereby loss of p53 causes a significant decrease in fertility particularly in female animals (Hu, Feng et al. 2007). Recently, Hu et al have shown that one crucial mechanism by which p53 regulates

maternal reproduction in mice is through its target gene leukemia inhibitory factor (LIF), a multifunctional cytokine, which plays a crucial role in blastocyst implantation in a p53-dependent manner (Hu, Feng et al. 2007; Hu, Feng et al. 2008; Hu 2009). Taken together, the observations made in invertebrates and mammals suggest that, besides its important role in maintaining genomic stability in somatic cells, p53 is also responsible to ensure faithful development, germ-line genomic integrity and fertility and thus the production of normal offspring that will survive and reproduce.

1.2. Low-frequency, highly penetrant human mutations

1.2.1. Introduction to p53 pathway low-frequency polymorphic variation

As mentioned above, evidence from over 30 years of intense research efforts has established that the p53 network of genes is crucial in suppressing cancer in vertebrates. In fact, it has been well documented that low-frequency, highly penetrant, inherited mutations in genes that are either directly in the p53 signaling pathway, or in closely interacting pathways, underlie many known cancer predisposition syndromes. These include Ataxia telangiectasia (caused by a homozygous mutations in the ATM gene) (Savitsky, Bar-Shira et al. 1995; Barlow, Hirotsune et al. 1996; Concannon 2002), Cowden disease, Lhermitte-Duclos disease, and Bannayan-Zonana syndrome (all caused by mutations in the PTEN gene) (Liaw, Marsh et al. 1997; Marsh, Dahia et al. 1997; Di Cristofano, Pesce et al. 1998), Xeroderma pigmentosum (caused by mutations in the XP genes) (Kraemer, Lee et al. 1994), Tuberous sclerosis complex 1 and 2 (caused by disruptions in the TSC1 and TSC2 genes) (Miyoshi, Nakau et al. 2002; Orlova and Crino

2010) or the Peutz-Jeghers Syndrome (caused by disruptions in the LKB1 gene) (Miyoshi, Nakau et al. 2002).

Many of the same somatic mutations of the p53 gene that are found at such high frequencies in cancer cells, can also be found as inherited mutations in individuals of the Li-Fraumeni cancer syndrome (LFS) (Varley, Evans et al. 1997; Varley 2003). This autosomal-dominant inherited cancer predisposition disorder with an estimated prevalence between 1:5000 – 1:20000 (0.02-0.005%) in the United Kingdom and the U.S.A (Lalloo, Varley et al. 2003; Gonzalez, Noltner et al. 2009) is characterized by a familial clustering of early onset tumors including osteosarcomas, soft-tissue sarcomas, breast cancers, leukemia and brain tumors (Li, Fraumeni et al. 1988; Malkin, Li et al. 1990). The causative role of mutant p53 in this syndrome has been supported by many mouse models. Mice that harbor germline mutations or deletions of the p53 gene display cancer phenotypes that closely resemble the human syndrome (Liu, McDonnell et al. 2000; Venkatachalam, Tyner et al. 2001; Lang, Iwakuma et al. 2004; Olive, Tuveson et al. 2004; Donehower and Lozano 2009; Lozano 2010). Specifically, animals with a germline deletion of one p53 allele ($p53^{+/-}$) or carrying hot-spot mutations in the DNA-binding domain of p53 (corresponding to the human R175H and R273H loci) develop a broad spectrum of tumors closely resembling LFS, particularly lymphomas, osteosarcomas, soft-tissue sarcomas and breast tumors significantly earlier than their wild-type counterparts (Liu, McDonnell et al. 2000; Venkatachalam, Tyner et al. 2001; Lang, Iwakuma et al. 2004; Olive, Tuveson et al. 2004; Donehower and Lozano 2009; Lozano 2010). Importantly, the median tumor incidence in the p53-targeting murine models is 18 months, and thus, for a mammal with a lifespan of 3 years, is similar to the

human tumor onset in p53-mutation carriers, whereby the average cancer incidence is of 50% by the age of 30 years (Malkin, Li et al. 1990; Donehower and Lozano 2009).

1.2.2. Genetic testing

Genetic screening for inherited p53 mutations and close surveillance of carriers in order to improve patient prognosis, has been a subject of great debate and is riddled with challenges (American_College_of_Medical_Genetics_Board_of_Directors 1995; Li 1995; Committee_on_Bioethics 2001; Evans, Lunt et al. 2010; Villani, Tabori et al. 2011). Some of the challenges have been: the wide range of tumors, the variability of the penetrance of the mutation, limited preventive options for most of the cancers and ethical issues concerning newborn screening (American_College_of_Medical_Genetics_Board_of_Directors 1995; Li 1995; Committee_on_Bioethics 2001; Achatz, Hainaut et al. 2009; Evans, Lunt et al. 2010; Villani, Tabori et al. 2011). However, females with a germline TP53 mutation can be given the option of prophylactic mastectomy to reduce the risk of breast cancer (Schneider and Garber 1993; Hartmann, Schaid et al. 1999; Thull and Vogel 2004). Furthermore, it has been recommended that radiation treatment in germline p53 mutation carriers should be assessed with scrutiny, as ionizing radiation has been shown to significantly increase the risk of second malignancies in this patient group (Hisada, Garber et al. 1998; Limacher, Frebourg et al. 2001; Varley 2003). A recent prospective observational study provides strong support for the utilization of genetic screening for inherited p53 mutations and subsequent presymptomatic surveillance of carriers in order to identify low-grade and premalignant cancers before they progress to a more malignant

state and, therefore, improve outcome (Villani, Tabori et al. 2011). In this study, asymptomatic p53 mutation carriers were surveilled using a protocol of non-invasive biochemical and imaging techniques. Of the thirty-three p53 mutation carriers identified, eighteen underwent surveillance. Remarkably, 3-year survival was 100% in the surveillance group, significantly higher compared to only 21% in the non-surveillance group.

Indeed, mass newborn screening for a relatively prevalent germline p53 mutation has been adopted in the state of Parana, Brazil (Achatz, Olivier et al. 2007; Achatz, Hainaut et al. 2009). This p53 mutation, R337H, is at an unusually high frequency (1:300 individuals, 0.3%) in the Brazilian population (Achatz, Olivier et al. 2007; Achatz, Hainaut et al. 2009). It associates with a 25% cancer risk at the age of 30 years compared to the estimated 50% by 30 years for the above-described p53 mutations. Although the genetic screening remains controversial due to the above-mentioned criticisms associated with genetic testing (Achatz, Hainaut et al. 2009), it is the first mass-implemented approach to utilize the inherited genetics of the p53 pathway to identify population groups at higher risk for cancer with the goal of early detection and improving the prognosis of cancer.

1.3. High-frequency, lesser penetrant human polymorphisms

1.3.1. Lessons from mice

Interestingly, results from multiple mouse models suggest that less penetrant alleles of p53 network genes could also significantly alter p53 signaling and affect cancer onset and progression (Mendrysa, McElwee et al. 2003; Mendrysa, O'Leary et al. 2006;

Alimonti, Carracedo et al. 2010; Lu 2010). For example, haploinsufficency for the murine double minute proteins MDM2 or MDM4 leads to increased p53 activity exhibited as increased sensitivity to DNA damage, decreased transformation potential and tumor development in mice (Terzian, Wang et al. 2007). Moreover, mouse models designed to express a hypomorphic MDM2 allele, which resulted in an approximately 30% reduction of overall MDM2 expression, had significantly increased p53 transcriptional activation and apoptotic activities (Mendrysa, McElwee et al. 2003). Furthermore, the mice also had a significantly reduced formation of cancer compared to wild-type MDM2 animals in an intestinal tumor model (Mendrysa, O'Leary et al. 2006). Similar observations have been recently made for the tumor suppressor PTEN, a key component of the PTEN/phosphatidylinositol 3-kinase (PI3K)/AKT pathway implicated in the regulation of multiple biological processes such as apoptosis, metabolism, cell proliferation and cell growth (Blanco-Aparicio, Renner et al. 2007). The PTEN/PI3K/AKT pathway is tightly integrated within the p53 network, and has been shown to affect the activity of p53 via multiple mechanisms (Mayo and Donner 2001; Zhou, Liao et al. 2001; Ogawara, Kishishita et al. 2002; Freeman, Li et al. 2003; Zhou, Gu et al. 2003; Li, Piluso et al. 2006). For example, PTEN has been reported to control p53 protein stability through antagonizing the AKT-MDM2 pathway (Mayo and Donner 2001; Zhou, Liao et al. 2001; Ogawara, Kishishita et al. 2002). PTEN has also been shown to regulate and guide p53 transactivation activity through a direct interaction that modifies its DNA binding affinity on endogenous p21 promoters (Freeman, Li et al. 2003; Li, Piluso et al. 2006). In a recent paper, Alimonti et al engineered a hypomorphic allele for murine PTEN that reduced PTEN expression by only 20% (Alimonti,

Carracedo et al. 2010). Interestingly, the authors demonstrate that animals expressing only 80% of normal PTEN levels show an increased susceptibility to developing various types of cancer (Alimonti, Carracedo et al. 2010). Together, these mouse models suggest that less penetrant alleles of p53 network genes could also significantly alter p53 signaling and affect cancer onset. Indeed, extensive study of human high-frequency genetic variants in the p53 network of genes supports this hypothesis.

1.3.2. High-frequency genetic variation

Human genetic variation, defined as the differences in DNA sequence within the genome of individuals, can be broadly classified into two different nucleotide composition classes, the single nucleotide variants (SNVs), which include single nucleotide polymorphisms (SNPs), point mutations and single nucleotide insertions or deletions (indels), and the structural variants (Feuk, Carson et al. 2006; Frazer, Murray et al. 2009). The latter group includes tandem repeats (short and variable number tandem repeats, also known as micro- and minisatellites), indels larger than 1 base-pair in size, copy number variations (CNVs) and other chromosomal rearrangements such as inversions and translocations, which constitute copy neutral variations, as well as copy neutral loss of heterozygosity (LOH) or homozygosity (Feuk, Carson et al. 2006). Together, these genetic variations encompass sequence changes in the genome with a size-range from single nucleotides (SNVs), over regions encompassing sizes between 2bp and 1kb in length (tandem repeats and indels), to regions spanning many kilo- or megabases (CNVs, inversions, translocations and LOHs) (Feuk, Carson et al. 2006).

One single nucleotide variant, the single nucleotide polymorphism (SNP), constitutes the most frequently studied form of high-frequency human genetic variation (Feuk, Carson et al. 2006; Frazer, Murray et al. 2009). SNPs can be distinguished from point mutations by their higher frequency in a population, whereby the minor allele frequency (MAF) of a SNP has to exceed the cutoff of at least 1% in any population (Feuk, Carson et al. 2006; Frazer, Murray et al. 2009). This cutoff, although commonly used in human genetics, is considered arbitrary, as all SNPs originate from germline mutations and rise in their frequency either by positive selective forces or randomly by genetic drift (Kelley and Swanson 2008). SNPs are usually bi-morphic, with two out of four possible nucleotide bases (i.e. Guanine, Cytosine, Adenine, Thymine) present at a particular allelic locus in a population (http://www.ncbi.nlm.nih.gov/snp). They arise from single historical mutational event in the human DNA, most frequently through physiological errors in DNA replication during cell division or alternatively induced by exogenous DNA-damaging agents (International_HapMap_Consortium 2003; Ku, Loy et al. 2010). These spontaneous or exogenously triggered mutational events occur in humans with a mutation rate of approximately 10^{-8} per site per generation (International_HapMap_Consortium 2003; Ku, Loy et al. 2010). The new mutation, or "allele", is termed the derived allele and, the former, the ancestral allele (International_HapMap_Consortium 2003).

Each allele is initially associated with a set of alleles of other SNPs that were present on the particular chromosomal background on which the SNP arose (International_HapMap_Consortium 2003; Kelley and Swanson 2008). The specific set of alleles observed on a single chromosome is called a haplotype

(International_HapMap_Consortium 2003; Kelley and Swanson 2008). The coinheritance of SNP alleles in haplotypes results in associations between the respective alleles in a given population that is called linkage disequilibrium (LD) (International_HapMap_Consortium 2003; Neale 2010). Over time, chromosomal recombination during meiosis results in the reduction of length of each haplotype, and thus the LD between SNPs, whereby the probability of recombination between two SNPs increases with the distance between them (Slatkin 2008). Therefore, in general, the linkage disequilibrium between SNPs declines with distance, as the association between the SNPs is reduced (International_HapMap_Consortium 2003; Slatkin 2008). Two measures of LD are commonly utilized, D' (D-prime) and r^2 (Rsquared) (Hill and Weir 1994; Devlin and Risch 1995; International_HapMap_Consortium 2003; Pittman, Myers et al. 2005). Both measures are based upon D, the basic pairwise disequilibrium coefficient, which describes the deviation of the observed from the expected frequency of a haplotype (Hill and Weir 1994; Devlin and Risch 1995; Pittman, Myers et al. 2005). Hereby, D' is a measure of the co-occurrence of the rarer allele with one of the possible alleles at the other marker, while the measure of r^2 has a more strict interpretation than that of D', as it also takes into account whether the marker loci have identical allele frequencies, i.e. whether all copies of the rarer allele occur exclusively with one of the possible alleles at the other marker (Hill and Weir 1994; Devlin and Risch 1995; Pittman, Myers et al. 2005). Both values are reported in the range of 0 to 1, whereby a value of 0 for both measures indicates an independent assortment of the alleles at different loci (linkage equilibrium), and a value of 1 suggests that the alleles at the different loci are perfectly linked (perfect LD) (Hill and Weir 1994; Devlin and Risch 1995). Therefore, if

two SNPs are in perfect LD with the r^2 measure, the allele at the one locus can always be predicted by the allele at the second locus (Hill and Weir 1994; Devlin and Risch 1995). The level of linkage disequilibrium is influenced by a number of factors, including the local rate of recombination, non-random mating, population subdivision and population bottlenecks, mutation, genetic drift and natural selection (Slatkin 2008).

1.3.3. Identification and cataloguing of SNPs

Historically, SNPs have been studied owing to their ability to alter the cutting sites of restriction enzymes and were thus sub-classified under the term restriction fragment length polymorphisms (RFLPs) (Donis-Keller, Green et al. 1987; Ku, Loy et al. 2010). They were one of the earliest genetic markers used in disease gene mapping (Ku, Loy et al. 2010). SNPs have been systematically catalogued since the inception of a genetic linkage map in 1987 (Donis-Keller, Green et al. 1987). Subsequently, a Single Nucleotide Polymorphism Database (dbSNP), a free public archive for genetic variation within and across different species, was developed in 1999 by the National Center for Biotechnology Information (NCBI) in collaboration with the National Human Genome Research Institute (NHGRI) (Sherry, Ward et al. 1999; Sherry, Ward et al. 2001). In addition, a map of human genome sequence variation containing 1.42 million SNPs has been launched by the International SNP Map Working Group in 2001, which provided an early resource for defining haplotype variation across the genome (Sachidanandam, Weissman et al. 2001). Facilitated in part by the publications of first complete reference DNA sequences of the human genome in 2001, which has contributed an important foundation for studying genetic variations among individuals in populations (Lander,

Linton et al. 2001; Venter, Adams et al. 2001), further refined maps of the human genomic variations have been published in the coming years, specifically by the three Phases of International HapMap Consortium (Phase I, II and III) and most recently by the 1000 Genomes Project (International_HapMap_Consortium 2003; International_HapMap_Consortium 2005; Frazer, Ballinger et al. 2007; Altshuler, Gibbs et al. 2010; Durbin, Abecasis et al. 2010).

The International HapMap Project is an international collaboration led by scientists in Japan, the U.K., Canada, China, Nigeria, and the U.S. (Altshuler, Gibbs et al. 2010). Utilizing a tag-SNP approach and complimentary sequencing, it has produced a data set comprising over 3.5 million common SNP genotypes of over 1184 individual samples from 11 population groups (e.g. Yoruba people in Nigeria, Japanese in Tokyo, Han Chinese in Beijing, and the U.S. residents with ancestry from Northern and Western Europe) and determined the frequencies of SNP alleles and genotypes for each population (International_HapMap_Consortium 2003; International_HapMap_Consortium 2005; Frazer, Ballinger et al. 2007; Altshuler, Gibbs et al. 2010). The 1000 Genomes Project is also an international research effort and aims to provide a deeper characterization of sequence variation between human genomes (Durbin, Abecasis et al. 2010). In Pilot 1, the project has utilized data obtained from low-coverage, whole-genome sequencing of 179 individuals from four different ethnic backgrounds, as well as high-coverage sequencing of 2 mother–father–child trios and exon-targeted sequencing of 697 individuals from seven different ethnic populations. To date, the project has identified approximately 15 million SNPs and determined the genotype frequencies in each population studied (Durbin, Abecasis et al. 2010). The HapMap, the 1000 Genomes

Project and the dbSNP databases are today the main genotype resources for researchers to map SNPs and identify the linkage disequilibrium between them. Together, there are currently more than 17 million catalogued SNPs, averaging about 1 every 200 nucleotides among the 3-billion nucleotide base pairs that constitute the genome of an individual ((Feuk, Carson et al. 2006; Frazer, Murray et al. 2009), http://www.ncbi.nlm.nih.gov/snp).

1.3.4. Phenotypic associations and potential functional impact of SNPs

These efforts to physically map and catalogue SNPs have been instrumental in leading to a notable advancement in the understanding of the biological functional roles of this form of genetic variation in a wide range of human traits and diseases, such as obesity, coronary heart disease, diabetes type II and cancer (Bond, Hirshfield et al. 2006; Iles 2008; Stadler, Thom et al. 2010; Wang, Li et al. 2011). Specifically, by allowing researches to utilize those genomic maps, they have enabled the design of association studies utilizing candidate gene, pathway-oriented and genome-wide approaches to identify SNPs, which associate with various forms of cancer, such as breast, ovarian, colorectal, bladder, prostate, head- and neck cancer, melanoma and soft-tissue sarcoma (Bond, Hu et al. 2004; Houlston, Webb et al. 2008; Thomas, Jacobs et al. 2008; Atwal, Kirchhoff et al. 2009; Thomas, Jacobs et al. 2009; Liu, Li et al. 2010; Bandele, Wang et al. 2011; Wang, Li et al. 2011). Moreover, together with the association studies, the physical maps of SNPs have enabled functional and molecular studies of polymorphisms, which contributed to the understanding of the biological functional roles of this form of genetic variation in human cancer. Specifically, SNPs have been found to reside in both

genic (exons, introns and genic promoters) and intragenic regions (Harris, Brill et al. 1986; Bond, Hu et al. 2004; MacPherson, Healey et al. 2004; Houlston, Webb et al. 2008; Thomas, Jacobs et al. 2008; Ernst, Kheradpour et al. 2011). Detailed molecular and functional analyses particularly of SNPs localized to genic regions have provided compelling evidence that SNPs can affect gene function through a wide range of mechanisms (Thomas, Kalita et al. 1999; Dumont, Leu et al. 2003; Bond, Hu et al. 2004; Zhang, Nicholls et al. 2010; Brest, Lapaquette et al. 2011; Ernst, Kheradpour et al. 2011). Non-synonymous exonic SNPs, which alter the protein sequence of a gene, have been shown to affect protein function (Thomas, Kalita et al. 1999; Dumont, Leu et al. 2003). Synonymous SNPs, which leave the protein sequence unchanged, or SNPs located in untranslated genic regions have been demonstrated to alter the binding affinity of microRNAs and thus affect the transcription efficiency of the gene (Wynendaele, Bohnke et al. 2010; Brest, Lapaquette et al. 2011). In intronic regions adjacent to alternative genic promoters, SNPs have been shown to affect transcription factor binding and thus directly result in changes in mRNA expression (Bond, Hu et al. 2004). Other intronic SNPs, located in the proximity of an exon-intron boundary have been shown to affect alternative splicing processes (Zhang, Nicholls et al. 2010).

1.4. The p53 pathway and high-frequency genetic variants in mice and humans

1.4.1. p53 codon72 polymorphism

Some of the most frequently studied SNPs that affect clinical phenotypes reside in the p53 network of genes (Harris, Brill et al. 1986; Bond, Hu et al. 2004). Hereby, two well-characterized SNPs in the p53 pathway are found in the p53 and MDM2 genes (p53

codon72 (Harris, Brill et al. 1986), rs1042522, C/G; MDM2 SNP309 (Bond, Hu et al. 2004), rs2279744, T/G) (Vazquez, Bond et al. 2008; Whibley, Pharoah et al. 2009; Grochola, Zeron-Medina et al. 2010). The different alleles of p53 codon72 encode either a proline (p53-codon72-Pro) or arginine (p53-codon72-Arg) residue and were first reported in 1988 by Buchman and colleagues (Buchman, Chumakov et al. 1988). The amino acid encoded by codon72 resides in a polyproline region of p53 that is located between the transactivation and the DNA binding domains (Walker and Levine 1996; Sakamuro, Sabbatini et al. 1997). Proline residues in this domain create binding sites for Src homology 3 (SH3) domain-containing proteins and have been shown to modulate signal transduction (Kay, Williamson et al 2000). This proline-rich region has been demonstrated to be important for p53 function, especially for its ability to induce apoptosis (Walker and Levine 1996; Sakamuro, Sabbatini et al. 1997).

The first study to provide evidence that the two different p53 isoforms encoded by the p53 codon72 SNP are not functionally equivalent was published in 1999 (Thomas, Kalita et al. 1999). For example, the authors observed in cell culture experiments that the codon72-Pro variant possesses increased ability to induce cell-cycle arrest through the transcriptional activation of the p21 (CDKN1A) gene (Thomas, Kalita et al. 1999). The authors reasoned that those allelic differences in transactivation activity might be due to the higher binding affinity of codon72-Pro to TAFII70 and TAFII32, two important factors of the basic transcriptional machinery (Thomas, Kalita et al. 1999). Subsequently, a higher capability of codon72-Pro to block a G1-S cell cycle transition has been suggested and the allele has been shown to have an increased propensity to express markers of cell cycle, as well as senescence, both under basal conditions and after

treatment with chemotherapeutics (Salvioli, Bonafe et al. 2005). Conversely, in the same, above-mentioned study that described an increased ability of codon72-Pro to induce cell cycle arrest, Thomas et al observed that p53-codon72-Arg was more efficient at both suppressing transformation by the oncogenes E7 and EJ-ras and initiating apoptosis (Thomas, Kalita et al. 1999). In a later report, Dumont and colleagues noted that p53-codon72-Arg has a stronger capacity to induce apoptosis (Dumont, Leu et al. 2003). Specifically, by employing a temperature-sensitive p53 mutant, they observed that p53-codon72-Arg is more efficiently translocated to the mitochondria, where it interacts with proapoptotic proteins such as GRP75 and Hsp60, thereby triggering apoptosis. Subsequently, three additional studies, utilizing various p53-inducible isogenic cell lines, also noted the greater apoptotic potential of p53-codon72-Arg both in the presence (Sullivan, Syed et al. 2004) and absence (Pim and Banks 2004; Bergamaschi, Samuels et al. 2006) of chemotherapeutics. Interestingly, one study presented data supporting a transcriptional-dependent mechanism to explain the different apoptotic potentials of the p53 isoforms. Specifically, Sullivan et al. studied H1299 lung carcinoma cell lines containing either inducible p53-codon72-Pro or p53-codon72-Arg and immortalized B-cell lines with the differing genotypes at this locus (Sullivan, Syed et al. 2004). In this report, the induction of p53-codon72-Arg and treatment with chemotherapeutics were shown to significantly induce up to 8-fold more apoptosis than p53-codon72-Pro and the chemotherapeutics. The authors demonstrated that this increased apoptosis associated with an elevated transcription of PUMA, PERP and AIP1, three known apoptotic effectors and targets of p53 transcriptional regulation, but not of two other p53 targets that are not direct effectors of apoptosis (MDM2 and p21). Similar findings were

observed in immortalized B-cell lines derived from individuals with wild-type p53 and differing genotypes at the codon72 locus (Sullivan, Syed et al. 2004). A subsequent report demonstrated that this difference in transcriptional regulation of downstream effectors was mediated, at least in part, by a higher binding affinity of p53-codon72-Pro to the protein iASPP (Bergamaschi, Samuels et al. 2006). Together with similar observations from other cell-based studies (Bonafe, Salvioli et al. 2002; Dumont, Leu et al. 2003; Pim and Banks 2004; Sullivan, Syed et al. 2004), these data support a model that in the context of wild-type p53, p53-codon72-Arg can be better at mediating the p53-dependent apoptotic response.

Intriguingly, multiple studies have demonstrated that the p53 codon72 SNP can affect apoptosis not only in the context of the wild-type p53 sequence but also in the context of p53 that has sustained somatic mutations. Interestingly, three reports suggest the mtp53-codon72-Pro isoform can associate with higher levels of apoptosis. In two studies, either Saos-2 cells (human osteogenic sarcoma) (Bergamaschi, Gasco et al. 2003) or H1299 cells (human lung cancer) (Vikhanskaya, Siddique et al. 2005) were stably transfected with a series of p53 mutants either proline (mtp53-codon72-Pro) or arginine (mtp53-codon72-Arg) for codon72. Together, both studies clearly show that in the context of some p53 mutations and the addition of certain chemotherapeutics, cells that express mtp53-codon72-Pro undergo more apoptosis than cells that express mtp53-codon72-Arg.

Two studies have proposed that several common mutants of mtp53-codon72-Arg bind with a greater affinity to a p53 family member, the tumor suppressor protein p73, thereby inhibiting its ability to induce apoptosis (Marin, Jost et al. 2000; Bergamaschi,

Gasco et al. 2003). For example, in an isogenic cell-based model, Marin and colleagues (Marin, Jost et al. 2000) demonstrated that conformational p53 mutants are capable of binding to and inactivating p73. This protein-protein interaction abrogates p73 ability to induce apoptosis by interfering with its DNA binding potential. Importantly, they observed that mtp53-codon72-Arg bound p73 with greater affinity than mtp53-codon72-Pro.

The precise molecular mechanism underlying these noted differences in apoptosis seems to merit further study (Vikhanskaya, Siddique et al. 2005), but the persistent observations that different alleles at codon72 collaborate with the p53 mutational status to influence a cell's ability to undergo apoptosis suggest the potential translational relevance of these studies. Specifically, knowing an individual's genotype at the p53 codon72 locus and the p53 mutational status of their tumor could help sub-classify the patient according to his/her probability to respond to certain chemotherapeutic agents in order to determine the best therapeutic strategy. Studies of 70 patients with inoperable, advanced head and neck squamous cell carcinomas, who all received cisplatin-based chemo-radiotherapy, support this hypothesis (Bergamaschi, Gasco et al. 2003; Sullivan, Syed et al. 2004). Specifically, those individuals who retained wild-type p53 in their tumors were found to associate with better response rates and overall and progression-free survival (Sullivan, Syed et al. 2004). However, in those patients that retained a wild-type p53, individuals with wtp53-codon72-Arg associated with the best response rates, and overall and progression-free survival, in concordance with its higher apoptotic potential. In those patients with mutant p53, individuals with mutp53-codon72-Pro associated with the best overall and progression-free survival, also in line with its relatively high apoptotic

potential compared to cells with mtp53-codon72-Arg (Bergamaschi, Gasco et al. 2003). Interestingly, those patients harboring the conformational p53 mutations that bind with greater affinity to p73 exhibited the greatest allelic differences.

Recent studies utilizing genetically engineered mice and mouse cells have shed light on the possible physiological effects of the p53 codon72 alleles (Phang and Sabapathy 2007; Reinbold, Luo et al. 2008; Zhu, Dolle et al. 2010; Frank, Leu et al. 2011). The first major challenge these studies have had to overcome is the fact that the region surrounding codon 72 is not highly conserved in mice and encodes an alanine residue. Zhu et al utilize mice with a humanized exon 4, which includes the sequence that encodes codon 72 (Arg/Pro) (Zhu, Dolle et al. 2010). Frank et al. utilize mice with humanized exons 4 to 9, which they refer to as humanized p53 knock-in gene (Hupki) (Frank, Leu et al. 2011). Zhu et al. observe that their mice carrying the Arginine residue associate with an increased rate of apoptosis in MEFs and in the small intestines after radiation compared to mice carrying the Proline residue (Zhu, Dolle et al. 2010). Frank et al. also observe that their mice carrying the Arginine residue associate with an increased rate of apoptosis in those tissues after radiation compared to their mice carrying the Proline residue (Frank, Leu et al. 2011). These data from both models are consistent with observations made in human tumor derived cell culture systems (Thomas, Kalita et al. 1999; Pim and Banks 2004; Sullivan, Syed et al. 2004; Bergamaschi, Samuels et al. 2006).

Frank et al. go on to show that the allelic differences in apoptotic potential for p53 codon72 are tissue-specific and that in another tissue it is the Proline isoform that associates with more apoptosis. Specifically, in thymocytes, p53 codon72-Pro mice have

increased apoptosis following ionizing radiation along with an increased transactivation of a subset of p53 target genes (Frank, Leu et al. 2011). The authors provide a possible explanation for the tissue specificity that is based on their observations that a group of genes induced by irradiation in the thymus are different than in other tissues, such as MEFs and the small intestines (Frank, Leu et al. 2011). This group of genes is made up of genes that have known roles in immunity and inflammation (Frank, Leu et al. 2011). Interestingly, the authors go on to explore the possible differences in the p53 isoforms directly in the inflammatory response. They were able to demonstrate that upon inflammatory stimulation of the mice through endotoxin lipopolysaccharide, those mice harboring the Proline residue are significantly more likely to succumb to septic shock than animals carrying the Argnine residue (Frank, Leu et al. 2011).

Intriguingly, both Frank et al and Zhu et al were unable to find any differences between the p53 isoforms in cancer risk or outcome. Frank et al crossed the proline and arginine mice with a mouse model that develops lymphomas (the Eμ-myc mouse), and with $p53^{+/-}$ mice that develop different cancers such as lymphomas and sarcomas (Donehower, Harvey et al. 1992; Schmitt, Wallace-Brodeur et al. 2000; Frank, Leu et al. 2011). They observed no significant difference in the survival or tumor spectrum between the mice harboring the different alleles (Frank, Leu et al. 2011). Zhu et al exposed their mice to chronic UVB treatment to induce skin cancer (Zhu, Dolle et al. 2010). However, they also observed no statistically significant differences in susceptibility to skin cancer development in the animals (Zhu, Dolle et al. 2010). Together, these data clearly demonstrate tissue-specific differences between the p53 Proline and Arginine isoforms in their abilities to induce apoptosis after DNA damage *in vivo*. Both models suggest that

the p53 codon72 SNP alone might not have a direct impact on cancer development. However, a trend is emerging in association studies in human cancer patients that this SNP might work in conjunction with another functional p53 pathway SNP in the MDM2 oncogene (MDM2 SNP309) to affect cancer risk (Wan, Wu et al. 2011).

1.4.2. MDM2 SNP309

MDM2 SNP309 is found at position 309 in the first intron of the MDM2 oncogene, which serves as a transcriptional enhancer region (Bond, Hu et al. 2004). Through computational analysis, the T to G allelic change introduced by SNP309 was predicted to increase the affinity of the Sp1 transcription factor by extending the length of an Sp1 binding site, which was subsequently validated through *in vitro* DNA-protein binding assays and reporter plasmid, cell-based assays (Bond, Hu et al. 2004). This model suggested that the G-allele of SNP309 would lead to increased transcription of MDM2, resulting in the attenuation of the p53 stress response and, potentially, tumor suppression. Experiments performed utilizing 43 tumor-derived cell lines of different SNP309 genotypes gave strong support to this hypothesis, particularly by demonstrating that cells G/G in genotype contain up to 8-fold higher levels of MDM2 RNA and, on average, 4-fold more MDM2 protein than cells T/T in genotype. Moreover, four other cell lines T/G in genotype for SNP309 were shown to contain intermediate protein levels (1.9-fold higher than T/T cells) (Bond, Hu et al. 2004). Similar trends have subsequently been observed in other studies utilizing tumor-derived cell lines (Bond, Hu et al. 2004;

Arva, Gopen et al. 2005; Hu, Feng et al. 2007; Leu, Wang et al. 2011), renal cell carcinomas (Hirata, Hinoda et al. 2007), normal esophageal tissues (Hong, Miao et al. 2005), and B-cell chronic lymphocytic leukemias (Gryshchenko, Hofbauer et al. 2008).

Together these observations support a model whereby Sp1 binds to the G-allele of MDM2 SNP309 with greater affinity than the T-allele, resulting in heightened MDM2 levels and the attenuation of the p53 pathway-mediated tumor suppression (Bond, Hu et al. 2004). Importantly, evidence from patient populations has also lent support to this model (Vazquez, Bond et al. 2008). For example, p53 mutation carriers (Li-Fraumeni individuals) with the G-allele of MDM2 SNP309 were shown to be diagnosed with tumors on average 7 years earlier than those that were T/T in genotype (Bond, Hu et al. 2004). This observation has been reproduced in three independent studies, in which p53 mutation carriers with the G-allele of MDM2 SNP309 were diagnosed with cancer on average 10, 12.5 and 16 years earlier than those who were homozygous for the T-allele (Bougeard, Baert-Desurmont et al. 2006; Ruijs, Schmidt et al. 2007; Marcel, Palmero et al. 2009). Earlier ages of onset associated with individuals with the G-allele, but no known p53 mutations, were also demonstrated in soft-tissue sarcomas (Bond, Hu et al. 2004), lymphoma (Bond, Hirshfield et al. 2006), leukemia (Swinney, Hsu et al. 2005; Phillips, Gerbing et al. 2010), head, neck and oral squamous cell carcinomas (Huang, Chen et al. 2008; Nakashima, Kondo et al. 2008), and cancers of the colon (Menin, Scaini et al. 2006), breast (Bond, Hirshfield et al. 2006; Wasielewski, Nagel et al. 2006), bladder (Sanchez-Carbayo, Socci et al. 2007), ovary (Bartel, Jung et al. 2008), brain (Khatri, Navaratne et al. 2008), melanoma (Firoz, Warycha et al. 2009) and liver (Yoon, Chang et al. 2008).

To date, a plethora of case-control studies have explored the association of MDM2 SNP309 with overall cancer risk (Hu, Jin et al. 2007; Wan, Wu et al. 2011). While many have shown a significant association of increased risk with the G-allele of SNP309 and various cancer types, such as lung, breast and endometrial cancer, others have failed to provide support for the observations (Hu, Jin et al. 2007; Ueda, Yamamoto et al. 2009; Fang, Yu et al. 2010; Joshi, Budhathoki et al. 2011; Wan, Wu et al. 2011). However, a recent meta-analysis of 66 papers that in total included over 25,000 cases and 30,000 controls comprising 30 tumor types provided evidence for an association of the G-allele of MDM2 SNP309 with increased overall cancer risk (Wan, Wu et al. 2011). The authors have shown that SNP309 G-allele carriers were associated with a significantly increased risk for various tumor types, such as breast, colorectal and lung cancer, with a combined odds ratio of up to 1.25 (95% confidence interval (CI)=1.13-1.37), providing supportive evidence to the model that MDM2 SNP309 serves as a low-penetrance tumor susceptibility marker (Wan, Wu et al. 2011).

Similar to p53 codon72, a recent study evaluated the effects of MDM2 SNP309 on the p53 pathway and cancer risk by creating a mouse model (Post, Quintas-Cardama et al. 2010). The authors were motivated by being able to explore its impact on cancer risk in a system that excluded the genetic and environmental heterogeneity of the cancer phenotype inherent to human association studies. To do this, Post et al generated mice that carry the humanized intron 1 of the MDM2 gene, containing either the G- or the T-allele of MDM2 SNP309 (Post, Quintas-Cardama et al. 2010). Interestingly, the authors find that cells from animals with a G/G genotype of MDM2 SNP309 have elevated MDM2 levels, reduced p53 levels, and decreased apoptosis. Moreover and in contrast to

the p53 codon72 mice, MDM2 SNP309 G/G mice have a shorter tumor latency and decreased survival, both in animals with two copies of wild-type p53 and in animals with one copy mutated ($p53^{515A/+}$) (Post, Quintas-Cardama et al. 2010). These data provide strong evidence that the G-allele of MDM2 SNP309 has a direct impact on cancer risk.

Interestingly, further studies in humans and human-derived material have demonstrated that the effects of the G-allele of MDM2 SNP309 on cancer can be modified by additional variables; such as gender, estrogen, and other p53 pathway SNPs (Alhopuro, Ylisaukko-Oja et al. 2005; Bond, Hirshfield et al. 2006; Bond, Menin et al. 2006; Lind, Zienolddiny et al. 2006; Bond and Levine 2007). Specifically, MDM2 SNP309 has been repeatedly shown to associate with allele- and gender specific differences in tumor diagnosis in various malignancies (Bond, Hirshfield et al. 2006; Bond, Menin et al. 2006; Bartel, Jung et al. 2008). Interestingly, female carriers of the G-allele have been shown to be diagnosed earlier in life with various cancers, such as colorectal cancer, diffuse large B-cell lymphoma, lung cancer and for highly estrogen receptor positive (>50% of tumor cells), but not for estrogen receptor negative, invasive ductal carcinoma of the breast (Bond, Hirshfield et al. 2006; Bond, Menin et al. 2006; Lind, Zienolddiny et al. 2006). This was shown to result in the enrichment of individuals with the G-allele in pre-menopausal women with these cancers, when compared to either post-menopausal women or men with the same cancers. Subsequently, similar trends have been observed in melanoma and osteosarcoma patients (Firoz, Warycha et al. 2009; Toffoli, Biason et al. 2009), while other studies suggest these trends will not be seen in every cancer and could be restricted to specific racial and ethnic backgrounds (Park, Choi et al. 2006; Bittenbring, Parisot et al. 2008). Recently, Hu et al. provided evidence for a

possible molecular mechanism for how the G-allele of SNP309 could accelerate tumor formation in this gender-specific and estrogen-dependent manner for some cancers, by demonstrating that the primarily female-specific hormone, estrogen, preferentially stimulated transcription of the MDM2 gene with the G-allele of SNP309 (Hu, Feng et al. 2007). Furthermore, a recent study provided evidence that the G-allele of SNP309 functions primarily in female carriers of germline p53 mutations to accelerate tumor formation (Atwal, Rabadan et al. 2008), which provided a potential genetic basis for a well-described sexual dimorphism in cancer risk observed in p53-mutation carriers (Chompret, Brugieres et al. 2000; Hwang, Lozano et al. 2003). This finding in conjunction with the above-described observations suggests that the G-allele of the MDM2 SNP309 locus could contribute to the increased cancer risk observed in female p53 mutation carriers through the preferential estrogen-dependent stimulation of transcription of the MDM2 gene (Hu, Feng et al. 2007). This model requires further testing, as estrogen signaling manipulation is being successfully used in both cancer prevention and treatment strategies (Jordan 2006) and the identification of a genetically defined population of women whose p53 pathway is attenuated in response to estrogen activation could positively affect clinical decisions.

A recent report has provided evidence that a lower-frequency SNP in the p53 network that resides in close proximity to SNP309 in intron 1 of the MDM2 gene can affect the G-allele of MDM2 SNP309 in its ability to interact with Sp1 and affect cancer risk (Knappskog, Bjornslett et al. 2011). Specifically, Knappskog et al report that the C-allele of MDM2 SNP285 (G/C) is found only in Caucasians and with a minor allele frequency of 3.6-4.1% compared to the 33.7-43.7% frequency of MDM2 SNP309. They

go on to show that the C-allele strongly reduces the binding of Sp1 to the G-allele of MDM2 SNP309 *in vitro* (Knappskog, Bjornslett et al. 2011). Moreover, the authors of the study show in a case-control study comprising 1993 ovarian, 1973 breast cancer patients and 3646 healthy controls that SNP285C reduces the risk of both ovarian (OR 0.74; CI 0.58–0.94) and breast cancers (OR 0.79; CI 0.62–1.00) among SNP309 G-allele carriers. They go on to identify a population of women G-in genotype for MDM2 SNP309 and G-in genotype for MDM2 SNP285 that has an increased risk for breast cancer (OR 1.27; CI 1.03–1.55) (Knappskog, Bjornslett et al. 2011).

Evidence for a possible interaction between a higher frequency SNP in the p53 network is provided by another recent publication (Wan, Wu et al. 2011). Specifically, the above-mentioned meta-analysis of MDM2 SNP309 case/control studies by Wan et al, comprising of 66 papers with over 25,000 cases and 30,000 controls suggests a significant interaction between MDM2 SNP309 and the p53 codon72 polymorphism (Wan, Wu et al. 2011). Hereby, the authors show that the combination of a SNP309G/G-codon72Pro/Pro genotype associated with an increased overall risk for cancer compared to the SNP309T/T-codon72Arg/Arg genotype (OR=3.38, 95% CI=1.77-6.47), suggesting that there is an association between MDM2 SNP309 and p53 codon72 regarding tumor susceptibility. Taken together, these reports suggest that in the future knowing the gender, estrogen exposures and the genotypes of interacting SNP loci, such as MDM2 SNP309 and SNP285, or SNP309 and p53 codon72, could help to more precisely predict an individual's cancer risk and could influence future cancer therapies.

1.4.3. p53 pathway haplotype structures

Most SNPs are thought to arise due to mutation and not influence the fitness of the organism and, therefore, increase in frequency in a given population by random genetic drift. Therefore, they achieve significant frequencies in the human population simply by chance (MacCallum and Hill 2006; Kelley and Swanson 2008). However, some mutational events will give rise to true functional alleles. These changes in function could confer a fitness advantage to individuals and therefore increase the frequency of the mutation in the population (positive selection). However, should the changes in function be detrimental to the organism, the frequency of the mutation will decrease (negative selection) (MacCallum and Hill 2006; Kelley and Swanson 2008).

One consequence of the positive selection of an allele of a SNP and the subsequent selective sweep in a population is the increase in frequency of neutral alleles in other SNPs closely linked to the selected allele, an effect called the hitchhiking effect. (MacCallum and Hill 2006; Kelley and Swanson 2008). This effect gives rise to hallmarks of allele-specific positive selection, namely, a reduction of nucleotide variation in the region of the genome surrounding the beneficial allele and a region characterized by a mix of long haplotypes harboring the selected allele and short haplotypes with the ancestral allele (MacCallum and Hill 2006; Kelley and Swanson 2008). Therefore, long high-frequency haplotypes indicate the action of positive selection, whereas non-selected haplotypes occur at lower frequencies and vary in length (MacCallum and Hill 2006; Kelley and Swanson 2008). Interestingly, Atwal et al determined that the frequency of the haplotype harboring the functional G-allele of MDM2 SNP309 deviates significantly

from the standard assumptions of models of selective neutrality by using multiple selection tests (Atwal, Bond et al. 2007).

Atwal et al. extensively analyzed the haplotype structure of MDM2 by combining genotype data generated from the International Hapmap Consortium, Celera Diagnostics and Sheba Cancer Research Center in Israel (Atwal, Bond et al. 2007). A low level of haplotypic variability was observed in Northern European populations and Ashkenazi Jewish populations, reflecting strong linkage disequilibrium across the whole gene. The level of linkage disequilibrium was perceived to be stronger than that due to a population bottleneck effect, arising from a founder population out of Africa. The effect of a founder population bottleneck effect would reduce the genetic variability of the entire genome and not just a single gene. By comparisons with extensive computer simulations of molecular evolution it was suggested that the G-allele of MDM2 SNP309 had experienced a positive selection sweep in human populations. The authors reasoned that an evolutionary selection pressure could act on the human p53 pathway (Atwal, Bond et al. 2007). As mentioned above, a growing body of evidence suggests that p53 regulates other human conditions that are under selective pressures, such as germ line maintenance, fertility and reproduction. Indeed, p53 codon72 polymorphism and MDM2 SNP309 have been shown to associate with recurrent implantation failure, missed abortion and *in vitro* fertilization treatment (Kay, Jeyendran et al. 2006; Atwal, Bond et al. 2007; Hu, Feng et al. 2007; Fang, Kong et al. 2009; Firouzabadi, Ghasemi et al. 2009; Kang, Feng et al. 2009).

The finding that the G-allele of MDM2 SNP309 had experienced a positive selection sweep gave rise to the hypothesis that functional SNPs in the p53 pathway

could be determined by haplotype studies that looked for evidence of natural selection. Recent work has shown that using signatures of natural selection to identify putatively functional sites and variants has been a fruitful exercise in the study of human cancer (Atwal, Kirchhoff et al. 2009). One of the first studies to demonstrate this was an investigation of the haplotype distribution of MDM4 (Atwal, Kirchhoff et al. 2009; Wade, Wang et al. 2010). The haplotype structure was analyzed in several human populations. Genotype data for SNPs were collected from 3 sources (i) a collection of 84 lymphoblastoid cell lines of African American and Caucasian ethnicity; (ii) the Hapmap Project repository including Caucasian individuals from the Centre d'Etude du Polymorphisme Humain collected in Utah, USA, with ancestry from northern and western Europe (CEU), and African (YRI) populations; and (iii) a cohort of 299 Ashkenazi Jewish controls from a breast cancer association study at Memorial Sloan Kettering Cancer Center. The low diversity of haplotypes from each dataset, resulting in atypical patterns of linkage disequilibrium, indicated the presence of candidate SNPs that may also modify the efficacy of the p53 pathway. In particular, one derived haplotype across the gene had increased to a high frequency (68% in the Caucasian lymphoblastoid dataset, compared to 30% in the African American population) in a relatively short amount of time in human history (Atwal, Kirchhoff et al. 2009). Subsequent association studies in 5 different patient populations revealed that these SNPs in MDM4 conferred an increased risk for, or early onset of, human breast and ovarian cancers in Ashkenazi Jewish and European cohorts, respectively (Atwal, Kirchhoff et al. 2009). This association was subsequently validated in two patient cohorts, whereby the MDM4 SNPs have been shown to associate with an accelerated age of onset of estrogen receptor

negative breast cancer (Kulkarni, Vazquez et al. 2009). These observations not only suggest that, like MDM2, MDM4 harbors SNPs that affect cancer in humans, but also other SNPs in the p53 network that affect cancer could also be under selection pressures.

1.4.4. Potential functional p53 pathway SNPs

The above-described studies suggest that the p53 stress response pathway could harbor more functional inherited genetic variants. Indeed, other less studied polymorphic variants have been reported for the p53 network genes CDKN1A, CDKN1B, p73, ATM and CASP8, which have shown repeatable associations with cancer predisposition and which we aim to describe in detail below (Grochola, Zeron-Medina et al. 2010).

Caspase-8 D302H

One important upstream effector of the extrinsic apoptotic pathway is caspase-8, which is activated after ligation of death receptors and initiates a cascade of downstream effector caspases (Bodmer, Holler et al. 2000; Hengartner 2000; Peter 2000). Recent evidence has shown that caspase-8 is also directly regulated by p53 under certain conditions such as after cellular exposure to cytotoxic drugs (Ehrhardt, Hacker et al. 2008). Recent studies have demonstrated a positive feedback loop between p53 and caspase-8 (Yao, Duan et al. 2007). The products of caspase-8 cleavage - death effector domains (DED) - can activate p53 to induce caspase-8 gene expression, which allows the continual replenishment of caspase-8 during apoptosis (Yao, Duan et al. 2007). Furthermore, in transfection experiments, p53 has been shown to be able to directly

regulate caspase-8 transcription through a p53-responsive sequence adjacent to the transcriptional start site of caspase-8 (Liedtke, Groger et al. 2003).

Recently, a single nucleotide polymorphism in the caspase-8 gene, termed D302H (rs1045485), has been reported to associate with altered breast cancer risk (MacPherson, Healey et al. 2004). This SNP results in an aspartic acid to histidine substitution at codon 302 of the gene, and is found at a high frequency in Europeans, with a minor allele frequency (MAF) of 20%, but has not been reported in Asian and African populations (http://www.ncbi.nlm.nih.gov). The first study noted an association of the minor histidine-encoding allele (H-allele) of D302H with reduced incidence of breast cancer in two independent cohorts in the United Kingdom (MacPherson, Healey et al. 2004). This finding was largely supported by two subsequent reports in a German population (Frank, Bermejo et al. 2005) and in a population in the United States (Sigurdson, Bhatti et al. 2007). Importantly, further supporting evidence of the protective role of the H-allele of SNP D302H in breast cancer was provided in a large case-control study comprising multiple ethnic cohorts of mostly European ancestry, carried out by the Breast Cancer Association Consortium (BCAC) (Cox, Dunning et al. 2007). This study consisted of an analysis of 14 independent studies with a total of 16,423 cases and 17,109 controls (Cox, Dunning et al. 2007). In line with these observations in case-control studies of sporadic breast cancer, a significant 11-years later age of breast carcinoma onset has been reported for the H-allele in a Spanish cohort of 390 patients carrying germ-line BRCA1/2 mutations (Palanca Suela, Esteban Cardenosa et al. 2009). Similar associations were seen in a study conducted in chronic lymphocytic leukemia (Enjuanes, Benavente et al. 2008). However, in contrast, no significant associations were seen in colon carcinoma (Pittman,

Broderick et al. 2008), and an inverse effect was reported for glioma, whereby the H-allele of SNP D302H seems to associate with an increased risk for this malignancy (Bethke, Sullivan et al. 2008).

p21 codon 31

The cyclin-dependent kinase (CDK) inhibitor p21 (CDKN1A) mediates the induction of cell cycle arrest in response to a variety of stimuli, mainly through its ability to inhibit both the kinase activity of CDK2 and CDK1 (Brugarolas, Chandrasekaran et al. 1995; Deng, Zhang et al. 1995; Macleod, Sherry et al. 1995; Abbas and Dutta 2009). The role of p21 in promoting DNA damage-induced G1 growth arrest relies to a great extent on its well-described transcriptional activation by p53 (Macleod, Sherry et al. 1995; Abbas and Dutta 2009). Given its crucial role in halting cellular proliferation, it is not surprising that p21 has been found to be frequently misregulated in human cancer, and the loss of its expression or function has been implicated in the genesis or progression of many human malignancies (Macleod, Sherry et al. 1995; Abbas and Dutta 2009). However, other studies also suggest that p21 can promote cancer, indicating a paradoxical effect leading to tumor-suppressing or tumor-promoting properties of p21, depending on the cellular context (Rowland and Peeper 2006; Abukhdeir and Park 2008; Abbas and Dutta 2009).

The non-synonymous codon 31 (C/A) SNP in the CDKN1A gene (rs1801270) results in an amino acid change from serine (p21-Ser) to arginine (p21-Arg) in a highly conserved region of the gene (Chedid, Michieli et al. 1994). Similar to p53 codon72 and MDM2 SNP309, the allelic frequency at this locus varies greatly between populations,

from a 4% prevalence of the Arg allele in a Swedish population to 50% in Chinese (Birgander, Sjalander et al. 1996). Functional studies seem to suggest that p21-Ser and p21-Arg variants share similar kinase inhibitory activity and growth suppression abilities (Sun, Hildesheim et al. 1995). However, the different genetic alleles that encode these variants have been shown to differ significantly in their transcriptional efficiency (Su, Sai et al. 2003; Johnson, Sherrington et al. 2009). Specifically, individuals who carry at least one p21-Arg-encoding allele associate with 38% lower p21 mRNA expression (Su, Sai et al. 2003). Similar observations were seen in a study designed to asses the role of the p21 codon 31 polymorphism in the ionizing radiation-induced, p53-dependent activation of p21 expression in chronic lymphocytic leukemia (Johnson, Sherrington et al. 2009). The authors noted that cell lines harboring at least one p21-Arg-encoding allele demonstrated an impaired p21 induction after radiation.

Possibly due to the seemingly conflicting reports linking p21 over-expression to both cancer suppressive as well as promoting effects, studies exploring associations of p21 codon 31 and cancer susceptibility report different cancer risk alleles. Specifically, some studies support the idea that the p21-Arg-encoding allele associates with increased risk for lung (Sjalander, Birgander et al. 1996), breast (Keshava, Frye et al. 2002), endometrial (Hachiya, Kuriaki et al. 1999), cervical (Harima, Sawada et al. 2001), bladder (Chen, Wu et al. 2002), head and neck (Rodrigues, Kawasaki-Oyama et al. 2003; Li, Liu et al. 2005) and prostate cancers (Huang, Wu et al. 2004). In contrast, others studies reported a decreased risk associated with the p21-Arg-encoding allele for cervical (Roh, Kim et al. 2001), esophageal (Wu, Wu et al. 2003) and endometrial cancers (Roh, Kim et al. 2004). In addition, some studies reported no clinical associations with this

locus (Su, Liu et al. 2003). Interestingly, a recent meta-analysis of 33 120 cancer cases and 44 954 controls from 49 publications with 66 individual case-control studies for the polymorphism, found a significant association of p21 codon 31 SNP with cancer risk in a tumor type-dependent manner in different ethnic populations (Liu, Li et al.). Specifically, the authors report that a significantly elevated risk for various tumor types was found among whites for the Arg-allele with an OR of up-to 1.48 (95% CI=1.20-1.83 for Arg/Arg vs Ser/Ser) (Liu, Li et al.). Among whites, when stratifying by cancer types, significantly increased risk was observed particularly for breast cancer (Liu, Li et al.). However, in Asian populations, significantly decreased risks were found for esophageal and gastric cancers (Liu, Li et al.). Together, the meta-analysis suggests that the p21 31Arg allele is a low-penetrant risk factor for cancer development, especially in white populations (Liu, Li et al.).

p27 -79C/T

p27 (CDKN1B), another important CDK inhibitor, monitors G_0 to S phase transitions by binding to and regulating the activity of various kinases involved in cell cycle progression (Abukhdeir and Park 2008; Chu, Hengst et al. 2008). Under conditions of cellular stresses, p27 integrates a variety of signal transduction pathways into a final proliferation checkpoint (Abukhdeir and Park 2008; Chu, Hengst et al. 2008). Mouse models suggest that p27 collaborates with p53 to modulate the cell cycle and suppress tumorigenesis in a tissue-specific manner (Philipp-Staheli, Kim et al. 2004; Damo, Snyder et al. 2005). For example, germline deletion of one or both alleles of p27 has been

shown to accelerate tumor development and associated mortality in p53 null mice, indicating potent synergy between loss of p27 and p53 (Philipp-Staheli, Kim et al. 2004).

The minor allele of a SNP within the p27 gene (-79C/T; rs34330), a C to T transition located 5' UTR at the site of nucleotide -79, was first proposed to confer an increased risk of prostate carcinoma in a hereditary prostate cancer (HPC) family-based study conducted at the Johns Hopkins Hospital in Baltimore, MD, USA (Chang, Zheng et al. 2004). Subsequently, the minor allele was also shown to associate with an increased risk of breast cancer in a case-control study, which included 368 Chinese breast cancer cases and 467 cancer-free controls (Ma, Jin et al. 2006). Furthermore, another study conducted in a large British population comprising 4470 cases of breast carcinoma and 4560 controls provided further evidence of a statistically significant association of the minor allele of p27 -79C/T with increased cancer risk (Driver, Song et al. 2008). Interestingly, a recent Spanish study demonstrated an increased risk for developing follicular papillary thyroid carcinoma associated with the minor T-allele of SNP p27 -79C/T in a study comprising 649 patients and 385 controls (Landa, Montero-Conde et al. 2010). The authors also show in a cell line panel that the T-allele associates with lowered p27 gene transcription and p27 protein levels, suggesting that allelic differences in gene transcription might affect the activity of p27 and subsequently cell cycle regulation, leading to allelic differences in cancer risk (Landa, Montero-Conde et al. 2010).

p73 G4C14-to-A4T14

p73 and p63 are two closely p53-related genes, which share a high level of sequence similarity with p53 and are capable of transactivating p53-responsive genes,

thereby promoting cell cycle arrest and apoptosis (Benard, Douc-Rasy et al. 2003; Murray-Zmijewski, Lane et al. 2006). Among the polymorphisms within the p73 gene, two common SNPs lie just upstream of the translational start site of exon 2 in so-called position 4 (G/A) and position 14 (C/T). These SNPs have been consistently shown to associate with allelic differences in risk for various cancers (Ryan, McManus et al. 2001; Li, Sturgis et al. 2004; Li, Wang et al. 2004; Hu, Miao et al. 2005; Niwa, Hirose et al. 2005; Pfeifer, Arbman et al. 2005; Chen, Sturgis et al. 2008; De Feo, Persiani et al. 2009). They are in complete linkage disequilibrium with each other and therefore the alleles are always inherited together to form the so-called p73 G4C14-to-A4T14 polymorphism (Kaghad, Bonnet et al. 1997; Ryan, McManus et al. 2001). The minor allele of G4C14-to-A4T14, A4T14, has been shown to associate with a significantly increased risk for squamous cell carcinoma of the head and neck (Li, Sturgis et al. 2004; Chen, Sturgis et al. 2008), gastric (De Feo, Persiani et al. 2009), colorectal (Lee, Hong et al. ; Pfeifer, Arbman et al. 2005) as well as endometrial cancers (Niwa, Hirose et al. 2005). Consistent with these findings, a study conducted in the United States with 1054 non-Hispanic whites diagnosed with lung cancer and 1139 cancer-free controls reported an association of the A4T14 allele with increased cancer risk (Li, Wang et al. 2004). Interestingly, a recent meta-analysis comprising 26 individual studies involving 8,148 cancer patients and 8,150 controls provides additional evidence that the A4T14 alleles associate with a significantly increased risk specifically of cervical, colorectal and head and neck cancers (Wang, Gao et al. 2011). In addition, in the stratified analysis of ethnicity, a significantly elevated cancer risk was found specifically in Caucasian populations (Wang, Gao et al. 2011). However, the A4T14 allele was shown to associate

with a decreased risk in lung cancer in a study comprised of 425 Chinese lung cancer patients and 588 cancer-free controls (Hu, Miao et al. 2005). Furthermore, an apparently protective effect of A4T14 was reported in a small study comprising 84 esophageal carcinoma patients and 152 controls in Ireland (Ryan, McManus et al. 2001). Interestingly, two independent studies performed in Korean populations showed no significant allelic differences for p73 G4C14-to-A4T14 in lung cancer risk (Choi, Kang et al. 2006; Jun, Park et al. 2007). Together, these observations suggest the p73 G4C14-to-A4T14 polymorphism is a low-penetrant risk factor for cancer development and its effects are dependent on the ethnic background of the population studied.

ATM Ser49Cys.

Another gene involved in the p53 stress response pathway is Ataxia telangiectasia mutated (ATM) (Kastan and Lim 2000; Levine, Hu et al. 2006). ATM is a serine/threonine-specific protein kinase that is activated by DNA double-strand breaks (Kastan and Lim 2000). It phosphorylates several key proteins that initiate the activation of the DNA damage checkpoint, leading to cell cycle arrest or apoptosis (Kastan and Lim 2000; Shiloh 2003). Current evidence suggests that p53 is the main target of ATM, which regulates the phosphorylation of p53 on Ser9, Ser15, Ser46 and Ser376, and subsequently impacts p53 transcriptional activity (Banin, Moyal et al. 1998; Canman, Lim et al. 1998; Waterman, Stavridi et al. 1998; Ashcroft, Kubbutat et al. 1999; Saito, Goodarzi et al. 2002). Furthermore, ATM also activates the checkpoint kinase CHK2 and MDM2, proteins that directly regulate the activity of p53 (Khosravi, Maya et al. 1999; Bartek, Falck et al. 2001; Maya, Balass et al. 2001).

A low-frequency SNP in the ATM gene, mainly present in populations with European ancestry, results in a serine (Ser-ATM) to cysteine (Cys-ATM) substitution in codon 49 (ATM Ser49Cys; rs1800054; MAF 3.3%; http://www.ncbi.nlm.nih.gov) (Buchholz, Weil et al. 2004). ATM Ser49Cys has been shown by multiple studies to harbor allelic differences in breast cancer risk (Buchholz, Weil et al. 2004; Stredrick, Garcia-Closas et al. 2006; Cox, Dunning et al. 2007). The Cys-ATM-encoding allele of ATM Ser49Cys has been shown to associate with an increased risk of bilateral breast carcinoma in a report comprising two large population-based studies from the U.S. and Poland including a total of 2856 breast cancer patients and 3344 cancer-free controls (Stredrick, Garcia-Closas et al. 2006). The results of a subsequent report suggest the increased risk associated with Cys-ATM is statistically significant, but only in a subset of progesterone receptor positive breast cancers (Cox, Dunning et al. 2007). The prospective Copenhagen City Heart Study genotyped ATM Ser49Cys in 10324 individuals (Dombernowsky, Weischer et al. 2008). Of those 10324, 2293 had developed various types of cancers. Interestingly, patients harboring Cys-ATM presented with an increased risk of melanoma, prostate carcinoma and cancers of the oral cavity/pharynx, but not the breast (Dombernowsky, Weischer et al. 2008). However, a stratification in patients according to the sex hormone receptor status was not performed, therefore a significant effect in the subset of patients with progesterone receptor-positive tumors cannot be excluded with certainty in this study.

1.4.5. The identification of novel functional p53 pathway SNPs

The more in depth study of these loci in the CDKN1A, CDKN1B, p73, ATM and CASP8 genes should help further define populations in their abilities to respond to stress, suppress tumor formation, and respond to DNA damaging therapies. The identification of other functional SNPs that mediate the p53 stress response will prove challenging, as there are over 50,000 SNPs in the NCBI SNP repository (dbSNP) in genes that have been implicated in mediating and regulating the p53 response (Vazquez, Bond et al. 2008). However, recently, approaches have been described that could help identify potential functional p53 pathway SNPs (Tomso, Inga et al. 2005; Noureddine, Menendez et al. 2009; Smirnov, Morley et al. 2009; Grochola, Zeron-Medina et al. 2010; Bandele, Wang et al. 2011). Specifically, these strategies identify SNPs, or groups of linked SNPs, that demonstrate allelic differences in characteristics similar, but not limited to, functional p53 pathway SNPs.

Smirnov et al. presented an approach with the potential to identify SNPs that affect the cellular DNA-damage response (Smirnov, Morley et al. 2009). In this DNA-damage response screen, the authors analyze individual variation in gene expression in response to ionizing radiation and conduct genetic linkage and association studies to map their regulators, testing for evidence of an association of their measurements with multiple polymorphic loci (Smirnov, Morley et al. 2009). Briefly, the authors used microarray experiments to measure the gene expression levels in immortalized B-cells from members of 15 CEPH pedigrees, made up of 30 individuals, 2 and 6 hours after exposure to 10 Gy of ionizing radiation. Subsequently, a computational genome-wide linkage analysis for each of the gene expression phenotypes was performed linking those

phenotypes to the genotypes of 4600 SNPs. As the p53 signaling pathway is important in the cellular responses to DNA-damage and stress, one would reason that p53 pathway SNPs could be identified in this screen. Indeed, the authors identified novel SNPs in three p53 pathway associated genes (CD44, FAS, PMAIP1), which were significantly associated with expression levels of their target genes (Smirnov, Morley et al. 2009).

A p53 pathway-oriented approach was designed to identify SNPs in potential p53 target genes that could affect the ability of p53 to regulate their transcription (Tomso, Inga et al. 2005; Noureddine, Menendez et al. 2009; Bandele, Wang et al. 2011). This approach first uses bioinformatics to scan for alleles of SNPs in p53 promoter response elements (REs) of potential p53 target genes that could modify p53 protein-DNA interactions (Tomso, Inga et al. 2005; Noureddine, Menendez et al. 2009). Subsequently, functional yeast- and mammalian cell-based assays are undertaken to assess the predicted effects of the alleles on p53-dependent reporter transactivation and gene expression (Tomso, Inga et al. 2005; Noureddine, Menendez et al. 2009). This method has been recently refined to include a multiplex format microsphere assay to further probe how alleles affect p53 ability to bind its RE (Noureddine, Menendez et al. 2009). Interestingly, this approach has identified over 30 p53 response elements in or adjacent to thirty-six genes (e.g. ADAR2, ARHGEF7, CASP1, DCC, EDN2, EOMES, RRM1, SCGB1D2, SEI1, TLR8), which harbor SNPs that significantly affect the cellular transactivation capabilities of p53 and its *in vitro* and *in vivo* binding characteristics in nuclear extracts (Tomso, Inga et al. 2005; Noureddine, Menendez et al. 2009; Bandele, Wang et al. 2011). The biological and clinical relevance of these findings remains to be determined.

1.5. Objectives of the study

In summary, the p53 pathway constitutes a major cellular gene network in vertebrates, which is crucial in directing the suppression of cancer formation, mediating the response to commonly used cancer therapies as well as the regulation of germ line maintenance, fertility and reproduction. It has been demonstrated that various cancer predisposition syndromes are caused by low-frequency, highly penetrant, inherited mutations in p53 network, the knowledge of which can be used to individually tailor cancer prevention and treatment strategies. Mounting evidence suggests that the p53 tumor suppressor pathway also harbors functional high-frequency, low-penetrance genetic variants that affect p53 signaling in cells, resulting in allelic differences in cancer risk, prognosis and responses to chemotherapy (Figure 1.1). Many genes that are important in the regulation of the p53 signaling network have been shown to harbor SNPs, which associate with repeatable allelic differences in human cancer. Moreover, the significant, combined effects of functional p53 pathway SNPs on cancer-related phenotypic manifestations could help to further define populations in their abilities to react to stress, suppress tumor formation and respond to therapies. Observations such as these, as well as the detailed functional and molecular analyses of p53 codon72 and MDM2 SNP309, suggest that the insight gained from a thorough analysis of the functional human genetics of this important network of genes could offer novel cancer treatment and prevention strategies as well as personalize those that already exist.

In this Thesis, we, therefore, first aim to identify and characterize other functional SNPs in this important cancer-signaling pathway. To do this, we take advantage of three well-characterized traits of known-functional p53 SNPs to mine genomic datasets,

namely signatures of recent natural selection, altered cellular chemosensitivities and gender- and age-dependent tumor incidence. Second, we aim to identify and functionally characterize the regulatory mechanisms of the identified polymorphisms.

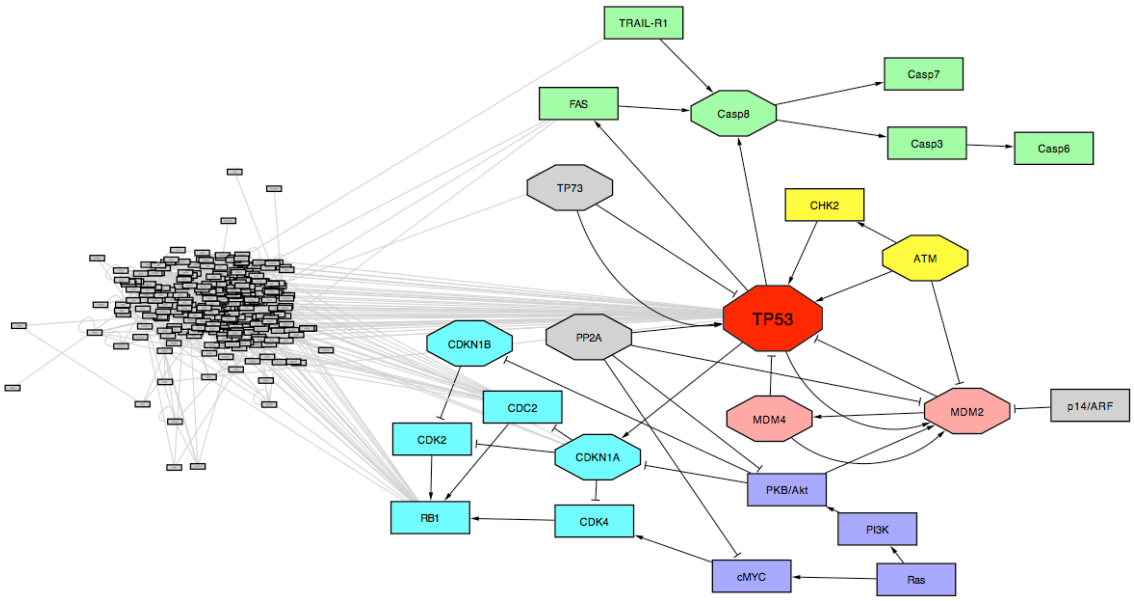


Figure 1.1. A map of selected p53 network genes that harbor known or potential functional SNPs (both in octagon-shaped boxes) and their closely interacting genes (rectangular boxes). The map depicts genes involved in the core regulation of p53 (light red boxes), the DNA-damage response (yellow), the extrinsic apoptotic response (green), the cell-cycle (turquoise) and oncogenic signaling (dark blue). The arrows (↓) show downstream activation and the inhibitory interactions are delineated by T-shaped (⊥) ends. A network map representing the p53 pathway is illustrated in the distant background.

II. Materials and Methods

2.1. Materials

2.1.1. Reagents

2.1.2. Antibodies, gene-specific assays, siRNAs

2.1.3. Cell lines

2.1.3.1. Primary cell lines grown in tissue culture

2.1.3.2. NCI60 human tumor cell line anticancer drug screen

2.2. Patients and controls

2.2.1. Soft-tissue sarcoma patients from Germany

2.2.2. Sarcoma patients from Australia

2.2.3. Ovarian carcinoma patients

2.2.4. Pancreatic ductal adenocarcinoma (PDAC) patients

2.2.5. Control cohort

2.3. Methods

2.3.1. Tissue culture

2.3.2. DNA analysis

2.3.3. Quantification of mRNA

2.3.4. Protein analysis

2.3.5. Analysis of DNA-protein interactions

2.3.6. Transfection-based assays

2.3.7. Data analysis

2.1. Materials

2.1.1. Reagents

All chemicals, unless otherwise stated, were obtained from Sigma-Aldrich (UK), Fisher Scientific (UK) or BDH Chemicals (UK).

Acrylamide gel for Electromobility Shift Assays (EMSA).

10% Acrylamide gel

10ml	40% acrylamide/bis 19:1 (Bio-Rad, Cat. No. 1610144)
4ml	5X TBE Buffer (Bio Rad, Cat. No. 161-0161)
26ml	ddH_2O
450µl	10% Ammonium Persulfate (APS)
25µl	Tetramethylethylenediamine (TEMED)

6% Acrylamide gel

8ml	30% acrylamide/bis 37.5:1 (Bio-Rad Cat. No. 1610158)
4ml	5X TBE buffer
28ml	ddH_2O
450µl	10% APS
26µl	TEMED

All acrylamide gels were pre-run for a minimum of 10 minutes in 0.5X TBE Buffer.

Agarose gel for DNA electrophoresis

Agarose powder (Fisher Scientific, UK) was weighted and dissolved in 1X TAE buffer (Bio-Rad, UK, Cat. No. 161-0743) at the appropriate concentration. The mixture was heated in a microwave oven to dissolve the agarose and the solution allowed to cool to 40°C. Ethidium Bromide was purchased from Bio-Rad (Cat. No. 161-0433) and added to a final concentration of 20μg/ml. The agarose solution was poured into a casting tray with the required comb and was left to solidify at room temperature.

Ammonium persulfate (APS)

10% (w/v) stock solution was prepared in distilled water and stored at -20°C in single-use aliquots.

Ampicillin stock

0.5g of the antibiotic was dissolved in 10ml of sterile distilled water, creating a 50mg/ml solution. This was stored at -20°C in single use aliquots.

Annealing buffer 2X

100mM	Tris pH7.6
10mM	Dithiothreitol (DTT)
20mM	$MgCl_2$
2mM	Spermidine

Appropriate amount of ddH_2O was added and the buffer stored at -20°C.

Black buffer 10X

400mM KCl

150mM HEPES (pH 7.9)

10mM Ethylenediaminetetraacetic acid (EDTA)

5mM DTT

50% Glycerol

Appropriate amount of ddH_2O was added and the buffer stored at -20°C.

Blocking solution

5% (w/v) fat-free milk (Marvel, UK) or 5% Bovine Serum Albumin (BSA; Sigma, Cat. No. A7906-100G) was prepared in 1X Phosphate Bufferd Saline (PBS) / 0.1% Tween.

Buffer 2 for nuclear extract preparation

2ml 0.5M HEPES pH7.9

150µl 1M $MgCl_2$

1ml 1M KCl

100µl 0.5M DTT, added just before use

96.75ml ddH_2O

Buffer 3 for nuclear extract preparation

4ml 0.5M HEPES pH7.9

8.4ml 5M NaCl

150µl 1M $MgCl_2$

100µl	0.2M EDTA
25ml	Glycerol
100µl	0.5M DTT, added just before use
62.25ml	ddH_2O

Complete Protease Inhibitor Cocktail

One Complete Protease Inhibitor tablet (Roche, Cat No. 11836145001) was dissolved in 2ml of sterilized distilled water as a 25X stock solution.

DNA marker

100bp and 1000kb DNA ladders were purchased from Anachem, UK (Cat. No. NM-7112, NM-7212) and Promega, USA (Cat. No. G210A, G571A) and were used as a size standard for agarose gel electrophoresis.

Elution buffer for Electromobility Shift Assays (EMSA)

40mM	KCl

The appropriate amount of ddH_2O was added

Freezing medium

90% (v/v)	Foetal Bovine Serum (FBS)
10% (v/v)	Sterile-filtered DMSO

Stored at -20°C.

Loading buffer for EMSA (10X)

2.4ml	1M Tris-Cl pH 6.8
3ml	20% sodium dodecyl sulphate (SDS)
3ml	Glycerol (100%)
1.6ml	Beta-mercaptoethanol
10ml	Bromophenol blue 0.006g

Phosphate Buffered Saline (PBS)

12.5 mM	NaCl
1 mM	Sodium dihydrogen phosphate, NaH_2PO_4
1.6 mM	Disodium dihydrogen phosphate, Na_2HPO_4

The appropriate amount of ddH_2O was added, the pH adjusted to 7.0 and the PBS autoclaved.

RIPA lysis buffer

150 mM	NaCl
1% (v/v)	NP40
0.1% (w/v)	SDS
50 mM	Tris-HCl, pH 8.0
1/25	Complete Protease Inhibitor Cocktail

The appropriate amount of ddH_2O was added.

6X SDS-PAGE sample buffer (Laemmli buffer)

750 mM	0.5M Tris-HCl pH, 6.8 (33.3ml in 50ml total)
30% (v/v)	Glycerol (15ml in 50ml total)
6% (w/v)	10% SDS (3g in 50ml total)
0.03% (w/v)	Bromophenol blue
60 mM	2-mercaptoethanol (210μl in 50ml total)

10X SDS-PAGE transfer buffer

725g	Glycine
145g	Tris

Final volume adjusted to 5L with distilled water. The buffer was then used with 20% ethanol as a 1X solution.

SDS solution

A 10% (w/v) solution of SDS was dissolved in water and stored at RT.

SDS-Polyacrylamide gels

4-15% gradient gels were purchased from Bio-Rad, UK (Cat. No. 456-1083)

Stripping buffer

100mM	2-Mercaptoethanol (1.4ml)
2%	SDS 10% (40ml)
62.5mM	1M Tris-HCl pH6.7 (12.5ml)

Final volume adjusted to 200ml with distilled water

5X TBE buffer

54g Tris Base

27.5g Boric acid

20ml 0.5M EDTA pH 8.0

Final volume adjusted to 1L with distilled water

Tris stock solution

Tris base was dissolved in water to provide a 1M solution, which was pH adjusted with concentrated HCl to the desired pH.

Primers used in PCR

All primers were designed with Primer 3 software (http://frodo.wi.mit.edu/primer3/) and Primer-BLAST (http://www.ncbi.nlm.nih.gov/tools/primer-blast/). The primers were synthesized by Operon, Germany and HPSF purified. All sequences are listed in the 5' – 3' direction:

ESR1, intron 1, forward (ChIP primer): TGAACCACCATGCTCAGTCT

ESR1, intron 1, reverse (ChIP primer): ACTCAGGGTCTCTGGGAAAC

ESR1, intron 1, forward (cloning primer): AACTGATATCCAGGGTTATGTGG

ESR1, intron 1, reverse (cloning primer): GCAAAACATGCACTCTCTGG

P21, promoter, forward (ChIP primer): GGGGCGGTTGTATATCAGG'

P21, promoter, reverse (ChIP primer): CGCTCTCTCACCTCCTCTGA

Oligonucleotides used in EMSA

All oligonucleotides were synthesized by Operon, Germany. The sequences are:

rs2234693 C-allele, sense: GGCCAAATGTCCCAGCCGTTTTATGCTT

rs2234693 C-allele, antisense: GGAAGCATAAAACGGCTGGGACATTTGG

rs2234693 T-allele, sense: GGCCAAATGTCCCAGCTGTTTTATGCTT

rs2234693 T-allele, antisense: GGAAGCATAAAACAGCTGGGACATTTGG

rs2234693 mutated E-Box, sense: GGCCAAATGTCCAAATTTTTTTATGCTT

rs2234693 mutated E-Box, antisense: GGAAGCATAAAAAAATTTGGACATTTGG

E2 E BOX, sense: GGGAGTGACATGAACAGGTGCACCCTGGCCTG

E2 E BOX, antisense: GGCAGGCCAGGGTGCACCTGTTCATGTCACTC

E2 E BOX, mutated E-Box, sense: GGGAGTGACATGAACCCGGGCACCCTGGCCTG

E2 E BOX, mutated E-Box, antisense:

GGCAGGCCAGGGTGCCCGGGTTCATGTCACTC

Water

DEPC-treated, nuclease-free water was purchased from Severn Biotech Ltd, UK (Cat. No. 20-9000-01) for PCR applications. Nanopure generated from the MilliQ water system was used for all other procedures.

2.1.2. Antibodies, gene-specific assays, siRNAs

Table 2.1. Primary Antibodies

Name	Antigen	Host	Type	Source (catalogue nr.)	Application	Working condition
Anti-TFAP4	TFAP4	Mouse	monoclonal	Abnova (H00007023-M01)	WB	1:1000
62A3	ER alpha	Mouse	monoclonal	Cell Signaling (2512S)	WB	1:1000
H-86	USF-1	Rabbit	polyclonal	Santa Cruz (sc-8983)	WB	1:200
N-17	TFAP4	Goat	polyclonal	Santa Cruz (sc-18593X)	ChIP EMSA	4ug 2ug
C-20	USF-1	Rabbit	polyclonal	Santa Cruz (sc-229X)	ChIP EMSA	4ug 2ug
6C5	GAPDH	Mouse	monoclonal	Milipore (MAB374)	WB	1:20000
F-10	ER alpha	Mouse	monoclonal	Santa Cruz (sc-8002X)	EMSA	2ug
C-4	ß-Actin	Mouse	monoclonal	Santa Cruz (sc-47778)	WB	1:1000

Table 2.2. Secondary Antibodies

Name	Antigen	Source (catalogue no.)	Application	Working condition
Anti-Rabbit	Goat	Bethyl (A120-101P)	WB ChIP	1:1000 4ug
Anti-Mouse	Rabbit	Dako (P0260)	WB ChIP	1:15000 4ug

Table 2.3. TaqMan Allelic Discrimination Assays for genotyping

Gene	SNP	TaqMan Allelic Discrimination Assay ID	Supplier
CCNG1	rs2069347	C___2000410_20	Applied Biosystems
CD44	rs187115	C_779820_10	
CSE1L	rs2426127	C__16230087_10	
ESR1	rs2234693	C_3163590_10	
KDR	rs2168945	C___1673863_10	
MDM2	rs2279744	Assay-on-demand	
MDM4	rs1563828	C_9493064_10	
MDM4	rs4245739	C_11623776_10	
PIAS1	rs1027154	C___1935268_20	
PPP2R2B	rs319227	C____803346_10	
PPP2R2B	rs319217	C___3065531_30	
PPP2R5E	rs11158491	C_2819093_10	
YWHAQ	rs6734469	C_29724290_10	

Table 2.4. TaqMan Gene Expression Assays

Gene	TaqMan Gene Expression Assay ID	Source
AR	Hs00907244_m1	Applied Biosystems
CD44	Hs00174139_m1	
CDKN1A	Hs00355782_m1	
ESR1	Hs01046818_m1	
GAPDH	Hs99999905_m1	
MDM4	Hs00159092_m1	
PPP2R5E	Hs00389197_m1	
USF1	Hs00273038_m1	
YWHAQ	Assay-on-demand*	

*Primer/probe sequences:
Primer forward: GTCCGAGCTGAGATCCATCTG
Primer reverse: CTCTGGATTAGTTGCATTGGCTATT
Probe: CGGTGCTGGAATTG

Table 2.5. TaqMan Allelic Discrimination Assays for allele-specific RT-PCR

Gene	SNP	TaqMan Allelic Discrimination Assay ID	Supplier
ESR1	rs2077647	C_11414978_10	Applied Biosystems
MDM4	rs4245739	C_11623776_10	

Table 2.6. siRNAs

Targeted gene	siRNA	Cat No	Supplier
TFAP4	siGenome SMARTpool	M-009504-00	Thermo Scientific Dharmacon
USF1	siGenome SMARTpool	M-003617-00	
Control	siGenome Control siRNA	D-001210-03-05	

2.1.3. Cell lines

2.1.3.1. Primary cell lines grown in tissue culture

Table 2.7. Cell lines

Name	Origin	Species
BT474	Breast cancer	Human
CAMA1	Breast cancer	Human
COS	Kidney	Monkey
HEK293	Embryonic kidney	Human
MCF7	Breast cancer	Human
T47D	Breast cancer	Human
U2OS	Osteosarcoma	Human
ZR75-1	Breast cancer	Human
HeLa	Cervical cancer	Human

2.1.3.2. NCI60 human tumor cell line anticancer drug screen

The National Cancer Institute 60 (NCI60) platform consists of 59 human tumor cell lines, representing 9 cancer types (Table 2.8): leukemia (represented by 6 cell lines), melanoma

(8 lines), cancers of the lung (9 non-small-cell lung cancer lines), colon (7 lines), central nervous system (6 lines), ovary (6 lines), breast (7 lines), prostate (2 lines) and kidney (8 lines) (Shoemaker 2006). The mutational status of p53 (Table 2.8, O'Connor, Jackman et al. 1997) and 20 important cancer-related genes, the genotypes of more than 100,000 SNPs (Affymetrix 125K chip), and the GI50 data for the NCI60 cell panel were obtained from the National Cancer Institute (NCI)/NIH Developmental Therapeutics Program. Microarray gene expression data of predicted transcription factors in the NCI60 cell line panel was obtained from the CellMiner™ (http://www.discover.nci.nih.gov/cellminer/) and the DTP (http://dtp.nci.nih.gov/mtargets/mt_index.html) databases. All data from both datasets included in the study was obtained from independent experiments, which were performed using the Affymetrix platforms U95, U95A, U133 and HU6800, as well as the Synteni (now Incyte, Inc., USA) cDNA array platform. The genomic DNA and mRNA from the NCI60 panel of cell lines was a generous contribution from the NCI-Division of Cancer Treatment and Diagnosis Repository Molecular Characterization Program.

Table 2.8. NCI60 cell line panel

Tissue of Origin	Cell line	AR mRNA levels		ESR1 mRNA levels		p53 mutational status*		
		ΔCt	Rank	ΔCt	Rank	p53 codon change	p53 base change	p53 amino acid change
Renal	786-O	19.11	37	21.47	40	278 CCT-GCT	C to G	278 P/A
Renal	A498	10.10	15	23.14	54	WT	WT	WT
Lung	A549/ATCC	8.88	8	22.60	48	WT	WT	WT
Renal	ACHN	22.35	51	16.46	20	WT	WT	WT
Breast	BT-549	8.51	5	22.97	52	NS	NS	NS
Renal	CAKI-1	12.22	23	15.15	13	WT	WT	WT
Leukemia	CCRF-CEM	21.13	45	19.87	32	248 CGG-CAG	G to A	248 R/Q
Colon	COLO 205	22.21	49	23.16	55	266 GGA-GAA	G to A	266 G/E
Prostate	DU-145	22.41	52	15.39	14	223 CCT-CTT	C to T	223 P/L
Lung	EKVX	17.51	34	19.09	29	del 187-224		del 187-224
Colon	HCC-2998	20.56	42	22.37	44	213 CGA-TGA	C to T	213 R/STOP
Colon	HCT-116	21.95	48	22.76	49	WT	WT	WT
Colon	HCT-15	22.76	55	22.50	45	wt/153 CCC-GCC	C to G	wt/153P/A
Leukemia	HL-60 (TB)	13.56	27	23.91	58	248 CGG-CTG	G to T	248 R/L
Lung	HOP-62	9.63	13	21.03	36	ins 212-225		ins 212-225
Lung	HOP-92	8.31	3	15.46	15	175 CGC-CTC	G to T	175 R/L
Breast	HS 578T	10.22	17	19.15	30	157 GAC-GAA	C to A	157 D/E
Colon	HT29	21.50	47	22.95	51	273 CGT-CAT	G to A	273 R/H
Ovary	IGROV1	20.83	43	22.58	46	WT	WT	WT
Leukemia	K-562	20.97	44	18.30	27	NS	NS	NS
Colon	KM12	22.69	54	23.56	57	179 CAT-CGT	A to G	179 H/R
Melanoma	LOX IMVI	23.39	58	22.59	47	WT	WT	WT
Melanoma	M14	13.52	26	19.36	31	266 GGA-GAA	G to A	266 G/E
Melanoma	MALME-3M	16.09	31	13.20	7	WT	WT	WT
Breast	MCF7	6.89	2	5.35	2	WT	WT	WT
Breast	MDA-MB-231/ATCC	17.65	35	21.06	38	280 AGA-AAA	G to A	280 R/K
Breast	MDA-MB-435	11.53	20	15.52	17	266 AGA-AAA	G to A	266 G/E
Leukemia	MOLT-4	9.29	12	22.09	43	WT	WT	WT
Breast	NCI/ADR-RES	11.55	21	17.89	26	del 126-132	G to A	del 126-132
Lung	NCI-H226	21.25	46	17.39	25	309 CCC-GCC	C to G	309 P/A
Lung	NCI-H23	11.62	22	20.70	35	246 ATG-ATC	G to C	246 M/I
Lung	NCI-H322M	9.93	14	13.36	8	248 CGG-CTG	G to T	248 R/L
Lung	NCI-H460	8.87	7	17.27	23	WT	WT	WT
Lung	NCI-H522	18.40	36	22.05	42	191 del G		191 del G
Ovary	OVCAR-3	22.34	50	17.24	22	248 CGG-CAG	G to A	248 R/Q
Ovary	OVCAR-4	20.26	39	13.91	11	WT	WT	WT
Ovary	OVCAR-5	10.91	18	21.05	37	ins 224		ins 224
Ovary	OVCAR-8	12.93	25	20.09	33	del 126-132		del 126-132
Prostate	PC-3	20.32	40	17.33	24	del 138		del 138
Leukemia	RPMI-8226	20.45	41	13.61	10	285 GAG-AAG	G to A	285 E/L
Renal	RXF-393	22.42	53	22.83	50	175 CGC-CAC	G to A	175 R/H
CNS	SF-268	12.30	24	15.92	19	273 CGT-CAT	G to A	273 R/H
CNS	SF-295	9.00	9	16.83	21	248 CGG-CAG	G to A	248 R/Q
CNS	SF-539	9.09	10	11.42	4	WT	WT	WT
Melanoma	SK-MEL-2	14.71	30	12.79	5	245 GGC-AGC	G to A	245 G/S
Melanoma	SK-MEL-28	10.96	19	13.59	9	145 TGT-GTT	T to G/G to T	145 C/V
Melanoma	SK-MEL-5	13.85	28	15.72	18	WT	WT	WT
Ovary	SK-OV-3	16.78	32	6.14	3	179 CAT-CGT	A to G	179 H/R
Renal	SN12C	19.13	38	20.19	34	336 GAG-TAG	G to T	336 E/STOP
CNS	SNB-19	8.41	4	23.49	56	273 CGT-CAT	G to A	273 R/H
CNS	SNB-75	8.62	6	18.85	28	258 GAA-AAA	G to A	258 E/K
Leukemia	SR	22.77	56	23.10	53	WT	WT	WT
Colon	SW-620	23.65	59	24.25	59	273 CGT-CAT	G to A	273 R/H
Breast	T-47D	6.52	1	4.92	1	194 CTT-TTT	C to T	194 L/F
Renal	TK-10	23.02	57	21.20	39	264 CTA-CGA	T to G	264 L/R
CNS	U251	9.10	11	21.59	41	273 CGT-CAT	G to A	273 R/H
Melanoma	UACC-257	13.87	29	12.86	6	WT	WT	WT
Melanoma	UACC-62	10.14	16	14.62	12	WT	WT	WT
Renal	UO-31	17.47	33	15.47	16	WT	WT	WT

Abbr. AR, Androgen receptor; ESR1, Estrogen receptor alpha; WT, wild-type, NS not specified

* O'Connor, Jackman et al. 1997

2.2. Patients and controls

2.2.1. Soft-tissue sarcoma patients from Germany

One hundred thirty patients (73 females and 57 males; ages 14–87 y; mean 56.3 y) diagnosed with soft-tissue sarcomas (STS) in the years 1991 to 2001 at the Surgical Clinic 1, University of Leipzig, Germany, and at the Institute of Pathology of the Martin-Luther-University Halle, Germany, were included in the study (Table 2.9). The mean observation time was 40.5 mo (range 2–198 mo). Sixty-three patients died from tumor-related causes within the observation time, whereas 67 patients were still alive at the time of follow-up. All patients underwent surgical treatment, and the subsequent mean survival time was 25.1 mo (range 2–119 mo). Eighty-three patients received postoperative radiotherapy and/or chemotherapy; 76 were treated with fractionated radiation (cumulative dose 60.4 Gy) and 26 received a combination treatment of doxorubicin and ifosfamide. The DNA of 66 patients was extracted from whole blood samples (n = 22) or from normal, pathologically confirmed tumor-free tissue, adjacent to the resection specimen (n = 44). The DNA of the remaining 64 patients was obtained from tissue within the confines of the tumor. Tumor tissue was available for 92 patients included in this study. Each person gave written and informed consent. Approval from the local ethics committee was obtained.

Table 2.9. Clinical and histopathological characteristics of the German STS cohort

	Number of Patients (%)
Stage	
I	14 (10.8%)
II	52 (40.0%)
III	34 (26.2%)
IV	30 (23.1%)
Grading*	
G1	18 (14.0%)
G2	72 (55.8%)
G3	39 (30.2%)
Tumor resection*	
Radical (R0)	96 (75.0%)
Not radical (R1)	18 (14.1%)
Not radical (R2)	14 (10.9%)
Tumor localization	
Head	2 (1.5%)
Neck	2 (1.5%)
Upper extremity	12 (9.2%)
Shoulder	1 (0.8%)
Thoracic wall	11 (8.5%)
Intrathoracal	0 (0%)
Intraabdominal	0 (0%)
Abdominal wall	12 (9.2%)
Retroperitoneum	15 (11.5%)
Hip/pelvis/gluteal	17 (13.1%)
Lower extremity	55 (42.3%)
Multiple localizations	3 (2.3%)
Tumor type	
Liposarcoma	30 (23.1%)
Fibrosarcoma	8 (6.2%)
Rhabdomyosarcoma	8 (6.2%)
Leiomyosarcoma	23 (17.7%)
Synovial sarcoma	9 (6.9%)
MFH**	33 (25.4%)
Neurogenic sarcoma	11 (8.5%)
Others	8 (6.2%)

*The grading of one patient and the resection status of two patients was not determined

**Malignant Fibrous Histiocytoma

2.2.2. Sarcoma patients from Australia

One hundred sixty-seven (72 females and 95 males; ages 9-80 y; mean 45.7 y) Australian sarcoma patients recruited for the International Sarcoma Kindred Study (ISKS) at the Peter MacCallum Cancer Centre, East Melbourne, Australia were included in the study. The breakdown into the tumor subtypes is given in Table 2.10. The ISKS was initiated in 2009 to investigate the prevalence and nature of heritable risk in sarcoma populations. The DNA of the patients was extracted from whole blood samples by the ISKS. Each person gave written and informed consent. Approval from the local ethics committee was obtained.

Table 2.10. Australian sarcoma cohort: breakdown into sarcoma subtypes

Sarcoma type	Number of Patients (%)
Soft-tissue sarcoma	
Liposarcoma	20 (21.1%)
Fibrosarcoma	20 (21.1%)
Rhabdomyosarcoma	4 (4.2%)
Leiomyosarcoma	20 (21.1%)
Synovial sarcoma	10 (10.5%)
Angiosarcoma	7 (7.4%)
Non otherwise specified (NOS)	10 (10.5%)
Others	4 (4.2%)
Total	95 (100%)
Primary bone sarcoma	
Ewing sarcoma	14 (29.2%)
Osteosarcoma	18 (37.5%)
Chondrosarcoma	16 (33.3%)
Total	48 (100%)
Non specified	
Total	24 (100%)
All sarcoma	167

2.2.3. Ovarian carcinoma patients

Ninety-four patients (ages 41–90 y; mean 64.8 y) diagnosed with ovarian cancer and known estrogen receptor status in the years 1997 to 2005 at the Institute of Pathology of the Martin-Luther-University, Halle, Germany, were included in the study (Table 2.11). DNA extraction from whole blood for genotyping was done at the Institute of Pathology. Each person gave written and informed consent. Approval from the local ethics committee was obtained.

Table 2.11. Clinical and histopathological data of the ovarian cancer cohort

	Total n
FIGO Stage	
I	27
II	7
III	56
IV	4
ER status	
high ER	40
low ER	12
no ER	42

Abbr.: FIGO, Fédération Internationale de Gynécologie et d'Obstétrique; ER, estrogen receptor

2.2.4. Pancreatic ductal adenocarcinoma (PDAC) patients

In the years 2005-2008, during my work as a residential trainee and postdoctoral researcher at the Department of General, Visceral and Transplantation Surgery (University of Ulm, Germany), I have helped recruit and establish a database of patients

who underwent primary surgery for PDAC at this institution. In total, the group comprised 107 patients (43 females and 64 males; ages 32-82 y; mean 62.1 y) who underwent primary surgery for PDAC in the years 2001-2007. The mean observation time was 16.2 mo (range 1-85 mo). Four patients died from non-tumor related causes. The clinical and histopathological data of the cohort are listed in Table 2.12. The genotypes of the SNPs included in this study were determined from blood samples or normal pathologically confirmed tumor-free tissue, adjacent to the pancreatic resection specimen. Additionally, during my time at the University Clinic of Ulm, I assisted in the determination of the p53 mutational status in the tumor tissue obtained during surgery for 90 patients (37 females and 53 males, Table 2.13). All persons included into the study gave a written and informed consent. Approval from the local ethics committee was obtained.

Table 2.12. Clinical and histopathological data of the PDAC cohort

	Total n
Stage	
I	7
II	21
III	67
IV	12
Tumour resection*	
Radical (R0)	77
Not radical (R1)	18
Not radical (R2)	11
Grading*	
G1	5
G2	48
G3	52
G4	1

* The grading and resection type of one tumor was not determined

Table 2.13. Mutations of p53 gene.

Mutation	n	Exon	Codon	Change	
				Nucleotide	Amino acid
Point mutations					
Missense	1	8	292	AAA to AGA	Lys to Arg
	1	10	338	TTC to CTC	Phe to Leu
	2	5	186	GAT to CAT	Asp to His
	1	9	322	CCA to CAA	Pro to Gln
	1	8	282	CGG to TGG	Arg to Trp
	1	3	28	TTG to TGG	Leu to Trp
	1	8	280	AGA to AAA	Arg to Lys
Silent	1	4	36	CCG to CCA	Pro
	4	6	213	CGC to CCC	Arg
	4	5	158	CGC to CGG	Arg

2.2.5. Control cohort

In collaboration with the German Red Cross Blood Transfusion Service NSTOB (Springe, Germany) we obtained blood samples of 499 blood donors (Germans with central European origins; 195 females and 304 males, (ages 19-68 y; mean 44 y) with no previous history of cancer (except in situ carcinoma after complete R0-resection). All persons included into the study gave a written and informed consent. Approval from the local ethics committee was obtained.

2.3. Methods

2.3.1. Tissue culture

Culture media. RPMI 1640 supplemented with L-Glutamine (Cat No. 21875-034) was purchased from Gibco-BRL, UK. Dulbecco's Modified Eagle's Medium supplemented with 4.5 g/L glucose and L-Glutamine (DMEM; Cat No. BE12-604F) was purchased

from Lonza, Belgium. FBS was purchased from PAA Laboratories, Austria. Penicillin/Streptomycin (P/S) was purchased from Gibco-BRL at 10,000 units/ml and used at a final concentration of 200 units/ml. All tissue culture dishes and flasks were obtained from Corning Inc (UK).

Maintenance of cell lines. All cell lines listed in Table 2.7 were cultured in complete medium (DMEM or RPMI 1640 , supplemented with P/S and 10% (v/v) FBS), in flasks or dishes maintained in a Binder incubator at 37°C in the presence of 5% CO_2. Medium was changed every 2-4 days depending on the cell lines. Upon reaching confluence, the cells were washed twice with 1X PBS and incubated with 1ml pre-warmed Trypsin-EDTA 0.05% (Gibco-BRL) at 37°C until the cells were detached from the flasks or dishes. Trypsin was inhibited by addition of the appropriate volume of fresh grown medium and this culture was then seeded onto the fresh flasks or dishes at the desired density. Cells in the exponential phase of growth were used for all analyses.

Freezing and thawing of cells. Cells were grown in 10cm plates to 70% confluence on the day of freezing. Cells from the growing culture were detached with 1ml trypsin-EDTA 0.05%, then resuspended in 9ml complete medium, pelleted and resuspended in 1 ml of freezing media at a concentration of 1-5 x 10^7 cells per ml. The cell suspension was transferred into 2ml freezing ampoules (Nunc). The vials were then labeled and cooled at the rate of 1°C per minutes in a Nalgen Cryo freezing container, in a -80°C freezer (New Brunswick Scientific) for at least 24 hr, before being transferred to a liquid nitrogen tank for long term storage. To thaw cells from the liquid nitrogen stack, vials were placed in a the 37°C water bath for 1-2 minutes and then transferred to a 10cm dish with 9ml of pre-

warmed fresh complete growth medium, and kept in the 37°C incubator overnight to recover.

2.3.2. DNA analysis

Purification of DNA. Genomic DNA of the patients and controls was extracted using the Innuprep Blood DNA mini-kit and the Innuprep DNA mini-kit by the AJ Innuscreen GmbH (Germany). Genomic DNA from cultured cells was extracted and purified using the DNeasy Blood and Tissue Kit (Qiagen, Cat. No. 69504) according to the manufacturer's instructions. In brief, cells were grown in 24- or 6-well formats using the appropriate medium until sub-confluent, harvested using trypsin, washed once with ice-cold 1X PBS, pelleted, resuspended in 200µl PBS and lysed with 20µl proteinase K (600mAU/ml) and 4µl RNAse A (100mg/ml). Buffering conditions were adjusted by adding the provided buffers and DNA selectively bound to the spin column membrane and eluted in the provided buffer. The concentration was determined by spectrophotometry using the Nanodrop spectrophotometer (Thermo Scientific, UK).

TaqMan Allelic Discrimination. The genotypes of all SNPs, except were indicated, were determined by PCR amplification and subsequent allelic discrimination using the appropriate genotyping assays (Table 2.3, Section 2.1.2). In brief, the allelic discrimination was performed as a multiplexed reaction in a 96-well format, using two primer/probe pairs in each reaction with a unique pair of fluorescent dye detectors to allow genotyping of the two possible variants at the SNP site in a target template sequence. Each well contained 2µl of DNA (conc. 10ng/µl), 12.5 µL of TaqMan

Genotyping PCR Master Mix (Applied Biosystems, Part Nr. 4371357), 0.625 µL of the genotyping assay, and 9.875 µL distilled water. Following PCR (Table 2.8), fluorescence was measured with the AB 7500 Real-Time PCR System (Applied Biosystems) and the genotype clusters scored using Sequence Detection Software (version 1.3; Applied Biosystems).

Table 2.14. Genotyping Thermal Profile

Stage	1	2	
Reps	1	40	
Temp (°C)	95	92	60
Time (min:sec)	10:00	00:15	01:00

Golden Gate Genotyping Platform. The Golden Gate platform (Illumina Inc, San Diego, CA, USA) is utilized in order to determine the genotypes of 672 SNPs in a high-accuracy, medium-throughput, 96-well plate assay. In brief, a multiplexed PCR reaction, which generates allele- and SNP-specific PCR products that are subsequently hybridized to beads is performed. Subsequently, the fluorescence on each bead is quantified resulting in a signal associated with a particular genotype (AA/AB/BB) at each locus and the genotype determined.

Gene copy number analysis. The copy number of the ESR1 gene was determined using a Pre-designed TaqMan Copy Number Assay (Applied Biosystems, Cat No. Hs04316604_cn) and RT-PCR based amplification on an AB 7500 Real-Time PCR System (Applied Biosystems). The assay consist of a FAM dye-labeled probe and unlabeled PCR primers. The assay is run simultaneously with a choice of TaqMan Copy Number Reference Assays (VIC dye-labeled probe) in a duplex real-time PCR. The copy number assay detects the target gene or genomic sequence of interest and the reference

assay detects a sequence that is known to be present in two copies in the diploid genome. Subsequently, relative quantitation analysis is performed by the CopyCaller software (Applied Biosystems).

PCR amplification and direct sequencing. The genotypes of MDM2 SNP309 in the control cohort were determined by PCR amplification and direct sequencing by the German Red Cross Blood Transfusion Service NSTOB (Springe, Germany).

Mutation analysis. The p53 gene mutation status of all 11 exons in the PDAC tumors was obtained by PCR amplification and consecutive GeneChip microarray analysis with the Genorama Genotyping Platform (Asper Biotech, Tartu, Estonia; AJ Innuscreen GmbH) in 66 patients (analysis performed until April 2008) and direct sequencing with the ABI 3730 DNA Sequencer (Applied Biosystems; AJ Innuscreen GmbH) in the subsequent 9 patients (analysis in May 2008). The somatic p53 mutational status of the STS tissues was determined by direct sequencing of exons 4 to 10 of the p53 gene by the Institute of Pathology of the Martin-Luther-University Halle, Germany as previously described (Taubert, Meye et al. 1998; Taubert, Wurl et al. 1998).

2.3.3. Quantification of mRNA

Preparation of cDNA. Cells were grown in a 24-well format in the appropriate medium according to the experimental procedure and the cDNA prepared from the mRNA using the FastLane cDNA extraction kit (Qiagen, USA; Cat. No. 215011) according to the manufacturer's instructions. In brief, the cell-culture medium was aspirated, the cells washed and lysed by adding the appropriate buffers. Next, the lysate was incubated in the

supplied gDNA Wipeout Buffer at 42°C for 5 minutes to effectively eliminate contaminating genomic DNA. 1µl of reverse transcriptase, 4µl buffer and 1µl primer mix were added to the lysate (template RNA) and the reverse transcription performed at 42°C for 30 min followed by an incubation at 95°C for 3 min to inactivate the reverse transcriptase. The cDNA was directly used for downstream applications or stored in aliquots at -80°C.

Real-time reverse-transcription PCR (RT-PCR). Real-time RT-PCR was performed as a relative quantification of synthesized cDNA using TaqMan gene expression assays on the AB 7500 Real-Time PCR System (Applied Biosystems) and the data analyzed using the Sequence Detection Software (version 1.3; Applied Biosystems). For the reaction, 2µl of 4-fold diluted cDNA was mixed with 10µl 2X TaqMan Gene Expression Master Mix, 1µl 20X TaqMan Gene Expression Assays (Table 2.4, Section 2.1.2) and ddH_2O added to a final volume of 20µl. The RT-PCR settings are shown in Table 2.15.

Table 2.15. RT-PCR Thermal Profile

Stage	1	2	3	
Reps	1	1	40	
Temp (°C)	50	95	95	60
Time (min:sec)	02:00	10:00	00:15	01:00

All values were normalized to the GAPDH expression (TaqMan Gene Expression Assay, Applied Biosystems) and the data reported as delta Ct values. All RT-PCR reactions, except the measurements of CD44, YWHAQ, PPP2R5E and androgen receptor mRNA levels were performed at the Ludwig Institute for Cancer Research in Oxford. The measurements of CD44, YWHAQ, PPP2R5E, androgen receptor and the corresponding

GAPDH mRNA levels (GAPDH cDNA Quantification Module, Roboscreen AG) in the NCI60 cell lines were performed by the Roboscreen GmbH, Germany with the ABI PRISM 7000 Sequence Detection System (Applied Biosystems).

Allele-specific quantitative real-time RT-PCR. Genomic DNA and mRNA were isolated from logarithmically growing cells. After reverse-transcription of the mRNA using the Fast Lane cDNA extraction kit (Qiagen), the cDNA was purified utilizing the NucAway Spin Columns (Ambion, Cat. No. AM10070). Subsequently, a duplex real-time PCR reaction (Table 2.16) was performed in a 96-well format, whereby two differently fluorescence-labeled probes specifically designed to target each allele of an exonic SNP (Table 2.5, Section 2.1.2) that is linked to the haplotype of interest in the used cell line. In the same duplex qRT-PCR reaction, using the extracted genomic DNA, a standard curve of the differences in fluorescence for both alleles over varying genomic DNA concentrations was generated and utilized to calculate the fold-difference in fluorescence for both alleles found in the cDNA synthesized from the RNA. Each well contained 2µl of cDNA, 12.5 µL of TaqMan Genotyping PCR Master Mix (Applied Biosystems, Part Nr. 4371357), 0.625 µL of the genotyping assay, and 9.875 µL distilled water. The AB 7500 Real-Time PCR System and the Sequence Detection Software (version 1.3; Applied Biosystems) was used; the RT-PCR are shown in Table 2.16.

Table 2.16. Allele-specific RT-PCR Thermal Profile

Stage	1	2	
Reps	1	45	
Temp (°C)	95	92	60
Time (min:sec)	10:00	00:15	01:00

2.3.4. Protein analysis

Protein extraction. Cells grown in monolayers in a 10 cm dish were washed two times with 1X PBS and 1 incubated with 1ml of pre-warmed Trypsin-EDTA 0.05% (Gibco-BRL) at 37°C, until the cells were detached from the dishes. Trypsin was inhibited by addition of the appropriate volume of fresh growth medium and the cells were pelleted at 1250 RPM at 4°C for 5 minutes. The pellet was washed twice in ice-cold PBS and re-pelleted. The pellet was then resuspended in the appropriate volume of RIPA lysis buffer (1000μl per 10 cm dish). The mixture was incubated on ice for 20 minutes with occasional vortexing and sonicated on ice 3X [30/120 sec on/off; 15% amp], incubated for an additional 10 minutes on ice with occasional vortexing and centrifuged at 10.000rpm at 4°C for 10 minutes. The resulting protein (supernatant) was removed to a fresh micro centrifuge tube and the cell debris discarded.

Determination of protein concentration. The protein concentration was estimated using a Bradford assay or the Pierce BCA Protein Assay Kit (Thermo Scientific, Cat. No. 23227). The absorbance was measured in duplicate for each sample at 595nm using a microplate reader (TECAN, sunrise microplate reader) and compared against a BSA standard curve.

SDS-polyacrylamide gel electrophoresis (PAGE). Equal amounts of protein (10-50μg) were loaded per lane with 5X loading buffer, boiled and loaded onto SDS-polyacrylamide gels. Gels were electrophoresed at 150V for 45 minutes in 1X Tris/Glycine/SDS running buffer (Bio-Rad, Cat. No. 161-0732) with prestained protein

markers (Bio-Rad, Cat. No. 161-0373), which were used as a size standard. The protein was then transferred onto nitrocellulose membrane (Bio-Rad, Germany, Cat. No. 162-0214) at 55V for 3h or 20V or O/N at 4°C. The membranes were stained with Ponceau S (Sigma Life Science, USA, Cat. No. P7170-1L) to confirm transfer and equal protein loading. After 2 wash steps with PBS/Tween, membranes were blocked with 5% milk in 1X PBS, containing 0.1 % Tween-20 (Bio-Rad, Cat. No. 161-0781). Primary antibody was dissolved in 5% milk (ERα antibody) or 1% milk (all other antibodies) and incubated O/N at 4°C (ERα antibody), or at RT for 1h (all other antibodies) or at RT or O/N at 4°C (all antibodies). After repeated washes in PBS/Tween, secondary antibody was added for 1h at RT. Membranes were washed at least 3 times for 15min with PBS-Tween and the results subsequently visualized by ECL detection (Enhanced Chemical Luminescence; Amersham Pharmacia Biotech, GE Healthcare, UK) using X-ray films (Amersham Pharmacia Biotech). For re-probing with another primary antibody, the membranes were incubated with the stripping buffer for 30 minutes at 50°C in a shacking incubator and subsequently extensively washed with PBS-Tween.

2.3.5. Analysis of DNA-protein interactions

Nuclear extract preparation. Cells were grown in monolayers in a 175 cm^2 flask until they reached 70-80% confluency and harvested by scraping in ice-cold PBS. The cells were spun down at 1000 rpm for 5min, washed, re-suspended in ice-cold PBS and re-pelleted. The pellet was re-suspended in the appropriate volume (200-300µl) of cold Buffer 2 and incubated on ice for 10min. After the cells have swollen, 0.1% final

concentration of NP-40 was added and the cells incubated for a further 5 minutes. After the cells have lysed, the lysate was centrifuged at 5000 rpm for 10 minutes. The supernatant was decanted and the pellet re-suspended in the appropriate volume (200-300µl) of cold Buffer 3 and incubated at 4°C under rotation for 30min. The lysate was then centrifuged at 10'000 rpm at 4°C for 10 minutes and the supernatant containing the nuclear extracts stored in aliquots at -80°C.

Expression of recombinant proteins. The proteins MyoD and Tal 1 were generated using the TNT Coupled Transcription/Translation Reticulocyte System (Promega, L5020). Full-length MyoD was expressed from the pRK5 expression plasmid using SP6 polymerase and Tal 1 was expressed from the pCITE2A *in vitro* expression vector, using T7 polymerase (generous gifts from S. De Val). To express recombinant protein, 25µl TNT lysate, 2µl reaction buffer, 1µl RNA polymerase (T7 or SP6), 0.5µl methionine amino acid mix, 0.5µl leucine amino acid mix and 1µg of DNA were mixed and the final volume adjusted to 50µl with distilled water. AP4 was exogenously expressed in HEK293 cells using an AP4-Flag-tagged pCMV vector (generous gift from C. Goding) and the Fugene-6 transfection reagent (Roche, Germany; Cat. No. 11 814 443 001). An empty pCMV vector served as a negative control.

Electromobility Shift Assay (EMSA). The annealing of the sense and antisense oligonucleotides was performed by mixing 5µl (1µg/µl) of each oligonucleotide (synthesized by Operon, Germany, section 2.1.1) with 10µl 2X anneal buffer and consecutive incubation at the following temperatures: 95°C, 65°C, 37°C and RT, for 10 minutes at each temperature step. For radioactive labeling, 1µl of the annealed oligonucleotide mix was incubated at RT for 30-60 minutes with 2.5µl Buffer 2 (New

England Biolabs, Cat. No. B7002S), 1µl 0.5mM dGTP, 1µl 0.5mM dATP and 1µl 0.5mM dTTP (Fisher Scientific, Cat. No BP2564-1), 0.5µl DNA Polymerase 1 Lg Klenow Fragment (New England Biolabs, Cat. No. M0210L), 17µl ddH_2O and 1µl of α-^{32}P-radioactively labeled dCTP (Perkin Elmer, Cat No. NEG513H250UC). For purification, 3ul of a 10X loading buffer was added to the EMSA probe and the mix run on a 10% non-denaturing acrylamide gel for 2 hours at 150V in 0.5X TBE buffer. Using a Perspex shield/box, an autoradiography film (GE Healthcare, UK) was placed directly on the gel for 10-20 seconds, the film developed and the gel area corresponding to the hot probe excised with a scalpel. 100µl of 40mM KCl was added to the gel slice containing the hot probe and incubated O/N at 37°C to elute. The sample was then spun down and the supernatant containing the purified hot probe stored. The radioactivity was counted using 1µl of the probe suspended in 1ml liquid scintillation cocktail (Fisher Scientific; Cat No. SC/6000/17) with the 1450 Microbeta Liquid Scintillation Counter (Perkin Elmer). For the binding reaction, 1-4µl recombinant protein from TNT reaction or 4µl nuclear extracts were incubated with 2µl 10X Black Binding buffer, 1-2µl poly dIdC (Roche, Cat. No. 10108812001), optionally 1µl of cold competing probe or 2µl antibody and ddH_2O added to a final volume of 20µl. The incubation time for the cold competing probe and the antibody was 10 and 20 minutes, respectively. Next, 2µl of the hot probe (diluted to 5000-20000 counts per minute/µl) was added and the reaction incubated for 20-30 minutes at RT. 2µl of 10X DNA loading buffer was added and the reaction loaded onto a 6% acrylamide gel and run at 150V in 0.5X TBE buffer for 2-3 hours at RT. The gel was dried at 80°C for 45-60 minutes on a gel drier on nitrocellulose filter paper and exposed for the appropriate time onto a radiography film or a phosphoimager screen

(Bio-Rad, Cat No. 170-7841). After exposure, the screen was scanned by the Molecular Imager Fx (Bio-Rad) and the data visualized and optionally quantified using the Quantity One 4.5.0 software (Bio-Rad).

Chromatin immunoprecipitation (ChIP). Chromatin immunoprecipitation was performed using the EZ-ChIP kit (Milipore, Cat. No. 17-371) according to the manufacturer's instructions. In brief, MCF7 cells were grown in monolayers in 20ml of appropriate medium until 80-90% confluency in a 15cm dish. Crosslinking was performed for 10 min using 550µl 37% formaldehyde (Sigma, Cat No. F8775) and the unreacted formaldehyde quenched with 2ml 10X Glycine. After 2 wash steps with ice-cold PBS, cells were scraped from each dish in 2ml PBS containing protease inhibitors (PI), centrifuged at 4°C and resuspended in SDS lysis buffer containing PI. 400µl of the lysate was aliquoted per tube. The sonication was performed on ice using the VCX 130 Sonics Vibra Cell sonicator equipped with a 3mm stepped microtip (Sonics & Materials, USA) with 4 sets of pulses [10/50 sec on/off] at 40% of the maximum amplitude (corresponding to 52 Watts). These shearing conditions were determined empirically to provide the desired shearing of crosslinked DNA to 200-1000 base pairs in length as confirmed on an agarose gel (maximum signal intensity at 400-500 bp). After centrifugation, Protein G Agarose beads were added to the diluted supernatant to remove proteins or DNA that may bind non-specifically to the Protein G agarose. After brief incubation and pelleting, 10µl (out of 500µl total suspension) of the supernatant was removed to serve as Input and saved at 4°C until the elution of the protein/DNA complexes. To the remaining supernatant, the antibody of interest (4µg; Tables 2.1 and 2.2) was added and incubated O/N at 4°C with rotation. Next, Protein G Agarose beads

were added to each immunoprecipitate (IP) and incubated for 1 hour at 4°C with rotation. The Protein G Agarose-antibody/chromatin complex (Tables 2.1 and 2.2) was consecutively washed in low, high and LiCl salt immune complex wash buffers, and twice in TE buffer. Next, each IP and the Input sample were eluted in the appropriate buffers. The eluate containing the anitbody-protein-DNA complex was reverse crosslinked and the freed DNA purified using buffers sand pin columns according to the manufacturer's instructions yielding 50µl eluted DNA.

Polymerase Chain Reaction (PCR) of chromatin-immunoprecipitated DNA.

PCR was run using 2-4µl template DNA, 5µl PCR Gold Buffer, 3µl $MgCl_2$ (final conc 3 nM), 4µl dNTP mix (10mM each dNTP; Bioline), 2.5µl forward and reverse primer (10pmol/µl each primer), 0.25-1 µl Amplitaq Gold Polymerase (Applied Biosystems; Part No. 4311820) and the final volume adjusted to 50µl with ddH_2O. The PCR conditions using the UnoCycler (VWR) are shown in Table 2.17. Next, 12.5µl of 5X loading buffer was added to each tube, the samples run on a 2% agarose gel containing ethidium bromide and the DNA bands visualized with UV light using the Benchtop UV Transilluminator (UVP BioDoc-It Imaging System; USA)

Table 2.17. PCR Thermal Profile

Stage	1	2		3
Reps	1	32-40		1
Temp (°C)	95	94	60	60
Time (min:sec)	10:00	00:30	01:00	10:00

Agarose gel electrophoresis. DNA samples were mixed with 5X DNA loading buffer (Bio-Rad, Cat. No. 161-0767; Bioline, at. Nr. BIO-37070) to a final concentration of 1X, before being run on an agarose gel of the appropriate percentage (1-2% gels), according to the size of the DNA bands to be visualized. The gels were made with 1X TAE buffer containing ethidium bromide and run at 90V with DNA markers. The ethidium bromide-stained DNA bands were then visualized using UV-radiation.

2.3.6. Transfection-based assays

siRNA-mediated gene expression silencing. The reverse transfection procedure was used to transfect siRNA into cells in 6-well and 24-well formats. For each well to be transfected in a 24-well format, 24pmol (40nM final concentration) of siRNA (Table 2.6, Section 2.1.2) was diluted in 100μl Opti-MEM I Medium (Invitrogen, Cat. No. 31985) without serum in the well of the tissue culture plate. After gentle mixing, 1μl Lipofectamine RNAiMAX (Invitrogen, Cat. No. 13778-075) was added to each well containing the siRNA molecules, the plate gently mixed and incubated for 10-20 minutes at RT. The cells to be transfected were diluted in the appropriate complete growth medium without antibiotics to a concentration of 10^5 cells per μl, 500μl of the diluted cells added to the Opti-MEM I Medium containing the Lipofectamine-siRNA complex and the plate gently mixed. The cells were incubated for 24-72 hours under standard conditions in the incubator.

Plasmid construction and cloning. A 465 base-pair long DNA sequence containing ESR1 SNP^{rs2234693} was PCR amplified from the blood donor DNAs homozygous for each

SNP allele (ID39, T/T; ID80 C/C) using sequence-specific primers (section 2.1.1) and the amplification of the PCR product confirmed on a 1% gel. The PCR products containing the amplified DNA sequence with either the T- or the C-allele were cloned into a pCR 2.1-TOPO plasmid vector using a TOPO TA Cloning kit (Cat. No. K4500-01) according to the manufacturer's instructions. This kit utilizes the TA cloning strategy, whereby the single, overhanging 3' deoxythymidine (T) residues of the linearized supplied vector are ligated by a viral Topoisomerase I to the complimentary single deoxyadenosine (A) residues at the 3' ends of PCR products generated by a non-template dependent terminal transferase activity of the Taq polymerase. The circularized vector was transformed into chemically competent E. coli cells and incubated on a selective LB agar plate containing 50μg/ml ampicillin for O/N incubation. Colonies carrying the plasmid were selected, grown in 4ml of LB medium containing 50μg/ml ampicillin O/N at 37°C. After incubation, 2ml of the LB medium was used to isolate plasmid DNA with the PureLink Quick Plasmid Miniprep Kit (Invitrogen, Cat No K-2100-10) according to the manufacturer's instructions, and all constructs subsequently verified by sequencing (Geneservice, Oxford, UK). The cells carrying the plasmid were grown in 400ml LB medium supplemented with ampicillin and the plasmids purified using the Qiagen Plasmid Maxi Kit (Qiagen, Cat No 12163) according to the manufacturer's instructions. Next, the pCR 2.1 TOPO vectors containing the sequences with each allele of SNP rs2234693 and a pGL3-TK-Promoter vector (Promega) were linearized with the restriction enzymes Xho I and Sst I using 1μg plasmid DNA, 2.5μl 10X React 2 Buffer (Invitrogen, Part No. Y92500), 1μl Xho I (Invitrogen, Cat. No. 15231-012), 1μl Sst I (Invitrogen, Cat. No. 15222-011), 15.5μl ddH_2O for 2 hours at 37°C and the reaction run

on a 1% acrylamide gel. The fragment corresponding to the vector was excised and the DNA extracted using the QIAquick Gel Extraction Kit (Qiagen, Cat No 28704) according to the manufacturer's instructions. The linearized plasmid was then re-circularized 16-24h at 14°C in the presence of the double-digested allele-specific sequence or annealed allele-specific oligonucleotides (Table 2.18) using 1U T4 DNA ligase (Invitrogen, Cat No 15224-041), 4µl 5X Ligase Reaction Buffer (Invitrogen, Cat No 15224-041), 190 fmol annealed oligonucleotides (Table 2.18), 32 fmol plasmid and distilled water adjusted to the final volume of 20µl. The ligated vector was then diluted 5-fold and transformed into chemically competent E. coli cells and incubated on a selective LB agar plate containing 50µg/ml ampicillin for O/N incubation. Colonies carrying the plasmid were selected, grown in 4ml of LB medium containing 50µg/ml ampicillin O/N at 37°C. After incubation, 2ml of the LB medium was used to isolate plasmid DNA with the PureLink Quick Plasmid Miniprep Kit (Invitrogen, Cat No K-2100-10) according to the manufacturer's instructions, and all constructs subsequently verified by sequencing (Geneservice, Oxford, UK). The cells carrying the plasmid were grown in 400ml LB medium supplemented with ampicillin and the plasmids purified using the Qiagen Plasmid Maxi Kit (Qiagen, Cat No 12163) according to the manufacturer's instructions. The plasmids were then transfected into cells and cultured in 24-well plates.

Table 2.18. Oligonucleotide sequences for plasmid construction

Allele, strand	Sequence
rs2234693 C-allele, sense	5' C(GTCCCAGCCGTTTTAT)$_3$C 3'
rs2234693 C-allele, anti-sense	5' TCGAG(ATAAAACGGCTGGGAC)$_3$GAGCT 3'
rs2234693 T-allele, sense	5' C(GTCCCAGCTGTTTTAT)$_3$C 3'
rs2234693 T-allele, anti-sense	5' TCGAG(ATAAAACAGCTGGGAC)$_3$GAGCT 3'

Transient plasmid transfections. Cells were grown in monolayers in a 24-well format or 10cm dishes until they reached 50-80% confluency. On the day of transfection the medium was changed to complete growth medium without antibiotics and transfected with the appropriate vector when in a log-growth phase. Fugene-6 transfection reagent and the plasmid vector were diluted with serum-free medium (without antibiotics) in a 3:1 ratio (i.e. 3μl Fugene-6 reagent per 1μg of plasmid) in an appropriate tube, the reaction mixed and incubated for 15-45 minutes at RT. The transfection reagent-DNA complex was added to the cells in a drop-wise manner and the wells swirled to ensure distribution over the entire plate surface. The cells were then incubated for 48 hours and harvested for extraction of recombinant protein or assayed using luciferase bioluminescence.

Luciferase Bioluminescence. Luciferase expression in plasmid-transfected cells was assayed using the Luciferase Assay System (Promega, USA, Cat. No. E1500). First, growth medium was removed from cells cultured in 24-well formats and the cells rinsed in ice-cold 1X PBS. 60 μl 1X Passive Lysis Buffer (Promega, USA; Cat. No. E1941) was dispensed into each culture well and incubated for 15-20 minutes on a rotating table. 20μl of the cell lysate was mixed with 50μl of the Luciferase Assay Reagent in a 96 well plate and the light produced measured using the Glomax Multi Detection System (Promega).

2.3.7. Data analysis

Haplotype analysis. Haplotype analysis was performed using publicly available data from the HapMap Project (HapMap Phase I, II and III, http://hapmap.ncbi.nlm.nih.gov/

(Altshuler, Gibbs et al. 2010)) and the 1000 Genomes Project Pilot 1 (http://www.broadinstitute.org/mpg/snap/ (Durbin, Abecasis et al. 2010)). All SNPs in high LD [$r^2 \geq 0.8$; MAF≥0.2; CEU] with a tag-SNP were included in the analysis.

Identification of conserved sequence motifs. The genomic databases (UCSC Genome Bioinformatics, http://genome.ucsc.edu/; VISTA http://genome.lbl.gov/vista/index.shtml) and multiple alignment computation analysis (CLUSTALW, http://align.genome.jp/; MULAN, http://mulan.dcode.org/) were used to compute the depth of conservation between the human, mouse, opossum and chicken genomes. All sequences were obtained from the NCBI assembly NCBI36/hg18 (http://www.ncbi.nlm.nih.gov/pubmed/)

Identification of functional elements in the human genome. The Encyclopedia of DNA Elements (ENCODE) project (http://genome.ucsc.edu/ENCODE/) (Birney, Stamatoyannopoulos et al. 2007) was used to access experimental data aiming to study chromatin accessibility, histone modification patterns and sequence-specific factors.

Transcription factor binding site analysis. Binding site analysis was performed using the computer algorithm ExPlain 3.0 and the Transfac 2009.3 database (Biobase; http://www.biobase-international.com/). A region of 25 base pairs up- and downstream of each SNP was included for each allelic variant in the analysis.

MicroRNA binding site analysis. Binding site analysis was performed using the computer algorithms SnipMir (http://www.microarray.fr:8080/merge/index?action =MISNP), RegRNA (http://regrna.mbc.nctu.edu.tw/) and Patrocles, (http://www.patrocles.org/). A region of 25 base pairs up- and downstream of each SNP was included for each allelic variant in the analysis.

Computer images. All autoradiographs were scanned using the Epson perfection 1660 Photo scanner and the Adobe Photoshop 8.0 software. Images were manipulated only as a whole for size, brightness and contrast. No signal was modified in relation to the whole image.

Statistical analysis. Survival analysis was performed using the Kaplan Meier analysis and the Cox's multivariate proportional hazards regression model with the SPSS 16.0-19.0 software (SPSS Inc., Chicago, IL, USA; IBM, USA). A Jonckheere-Terpstra, Wilcoxon Test, Mann-Whitney, T-Test, and a permutation test were applied to determine the statistical significance of the noted increase of the median or mean age of tumor diagnosis. A Bayesian estimate, a binomial test, the Chi-square test and the Fisher's exact tests are used to compare the frequency distributions of the genotypes. Case and controls were matched for both gender and age. Correlation analysis was carried out using the Spearman-Rho and Pearson tests. The expression of mRNA levels in cell culture experiments was compared using all values from each measurement and experiment with a two-sided T-test or the One-way Analysis of Variance (ANOVA) Test.

III. Results

3.1. Recent natural selection identifies a genetic variant in a regulatory subunit of protein phosphatase 2A that associates with altered cancer risk and survival

3.1.1. Introduction

3.1.2. Results

3.1.2.1. A polymorphism in the PPP2R5E gene resides in an evolutionary selected genomic region

3.1.2.1. A selected SNP in the PPP2R5E gene associates with altered sarcoma risk and survival

3.1.3. Discussion

3.1.1. Introduction

The SNPs in the p53 pathway that are most frequently studied are found in the p53 and MDM2 genes (p53 codon72; (Harris, Brill et al. 1986), MDM2 SNP309; (Bond, Hu et al. 2004)). The allelic frequencies for both SNPs vary significantly in different ethnic and racial populations. For example, the G allele of MDM2 SNP309 was found to be at 10% in African Americans (Atwal, Bond et al. 2007), 33% in Northern Europeans (Bond, Hirshfield et al. 2006), 45% in Asians (Hong, Miao et al. 2005), and 50% in Ashkenazi Jewish individuals (Bond, Hirshfield et al. 2006). To explain the significant differences for both polymorphisms, allele-specific selective advantages and disadvantages for evolution and/or reproduction have been proposed ((Chapter 1.4.3; (Beckman, Birgander et al. 1994; Atwal, Bond et al. 2007; Hu, Feng et al. 2007; Shi, Tan et al. 2009)). Indeed, two recent reports have provided evidence of recent natural selection for both polymorphic loci (Atwal, Bond et al. 2007; Shi, Tan et al. 2009)). For example, in the first report the MDM2 haplotype structure was determined in Caucasians (Atwal, Bond et al. 2007). It was noted that the unusually long and frequent haplotype that harbors the G-allele of MDM2 SNP309 deviates significantly from the standard assumptions of selective neutrality using multiple selection tests. These include an entropy-based selection test that compares both the frequency and long-range correlations of an allele with a simulated model in which the allele is selectively neutral. Although the precise reason(s) for the selection pressure on these functional p53 pathway SNPs is still under investigation, many reported observations suggest that selection pressure on the p53 pathway may be possible. For example, a well-regulated p53 pathway has been

shown to be crucial not only for murine tumor suppression and the response to DNA damaging therapies, but also for proper embryonic development, embryonic implantation, pregnancy rates, and litter sizes ((Chapter 1.1.3; (Lozano and Zambetti 2005; Atwal, Bond et al. 2007; Hu, Feng et al. 2007)). Indeed, allelic differences have been noted for both functional p53 pathway SNPs in human fertility (Kay, Jeyendran et al. 2006; Fang, Kong et al. 2009; Firouzabadi, Ghasemi et al. 2009; Kang, Feng et al. 2009). However, regardless of the precise reasons for the selection pressures, signs of natural selection of p53 variants seem to be a characteristic of functional p53 pathway SNPs as has been further validated by a recent study of the genetic variation of the MDM4 oncogene ((Chapter 1.4.3; (Atwal, Kirchhoff et al. 2009)).

The identification and description of high-frequency SNPs in the p53 stress response pathway that alter cancer risk, progression, and outcome will not only potentially increase our understanding of this tumor suppressor pathway in humans, but could also lead to their use in personalizing prevention and therapeutic strategies. Therefore, in this Chapter, we attempt to accelerate the identification of functional genetic variants, by incorporating the above-described characteristics of functional p53 pathway SNPs into an analysis of 142 genes known to be important in mediating and regulating the p53 stress response. Specifically, genetic variants are sought in those genes, which associate with haplotypes that show signs of natural selection, and differences in STS risk and survival. The work presented herein is a result of the collaboration with Alexei Vazquez from The Institute for Advanced Study in Princeton, New Jersey, USA and has been published during the continued work on my Thesis (Grochola, Vazquez et al. 2009).

3.1.2. Results

3.1.2.1. A polymorphism in the PPP2R5E gene resides in an evolutionary selected genomic region

To identify genetic variants in p53 stress response genes that deviate from the standard assumptions of selective neutrality, we utilized a previously described and publicly available map of recent positive selection of the human genome (Voight, Kudaravalli et al. 2006). In Haplotter (http://hg-wen.uchicago.edu/selection/haplotter.htm), recent positive selection is determined using a haplotype-based approach that looks for enrichment of the classic signal for strong directional selection using the phase II data of the HapMap. The classic signal for strong directional selection is an extended haplotype with very low diversity, or high homozygosity. This is thought to occur when a selected allele increases in frequency so quickly that insufficient time remains to allow recombination to reduce its length. Haplotter utilizes a test statistic called the integrated haplotype score (iHS), which is a measure that includes the degree of haplotype homozygosity around a given SNP (Sabeti, Reich et al. 2002). Highly positive and negative iHS scores (>2.5 or <−2.5) denote SNPs that harbor higher haplotype homozygosity, compared with other SNPs with similar allele frequencies in the genome. A large negative score denotes an unusually long haplotype carrying the derived (nonancestral) allele; a large positive score denotes a long haplotype carrying the ancestral allele.

Importantly, a genomic region or gene is only noted to be positively selected when it contains numerous SNPs with extreme iHS scores (Voight, Kudaravalli et al.

2006). An empirical p-value is assigned to each gene as a fraction of genes in the genome with the same number of SNPs with extreme iHS scores, in a window of 50 consecutive SNPs, that is equal to or larger than the given gene. We analyze 142 genes, which have been shown to be important in mediating and regulating the p53 stress response (Table 3.1.1). Those genes were carefully selected based on entries in the Kyoto Encyclopedia of Genes and Genomes (KEGG) Pathway Database (Kanehisa 1997) and complemented by a comprehensive literature search. The KEGG Database is a publicly available and widely utilized resource, which records published networks of molecular interactions in cells (Kanehisa 1997). It is curated by the Kyoto University and the University of Tokyo and funded by the Japanese Ministry of Education (Kanehisa 1997). The analysis of the 142 p53 stress response genes shows that only one gene resides in a genomic region that shows significant evidence of recent positive selection. A gene that encodes a regulatory subunit of the protein phosphatase 2A (PP2A) holoenzyme, PPP2R5E (protein phosphatase 2, regulatory subunit B′, ε-isoform), is found to be in a region of the genome that shows significant signs of natural selection in the CEU population of the HapMap ($p=0.006713$). Specifically, out of the 122 genotyped SNPs in PPP2R5E, with minor allele frequencies ≥10%, 29 SNPs have extreme iHS scores (Table 3.1.2, 26 SNPs with iHS >2.5 and 3 SNPs with iHS <-2.5).

Table 3.1.1. A list of 142 genes important in the p53 stress response and included into the genomic scan for natural selection.

Gene Symbol	Description
AKT1	v-akt murine thymoma viral oncogene homolog 1
AKT2	v-akt murine thymoma viral oncogene homolog 2
APAF1	apoptotic peptidase activating factor 1
APC	adenomatous polyposis coli
APEX1	APEX nuclease (multifunctional DNA repair enzyme) 1
ATM	ataxia telangiectasia mutated
ATR	ataxia telangiectasia and Rad3 related
AURKA	aurora kinase A
BAI1	brain-specific angiogenesis inhibitor 1
BAX	BCL2-associated X protein
BBC3	BCL2 binding component 3
CCNG1	cyclin G1
CD82	CD82 molecule
CDK2	cyclin-dependent kinase 2
CDK8	cyclin-dependent kinase 8
CDKN1A	cyclin-dependent kinase inhibitor 1A (p21, Cip1)
CDKN1B	cyclin-dependent kinase inhibitor 1B (p27, Kip1)
CDKN2A	cyclin-dependent kinase inhibitor 2A (melanoma, p16, inhibits CDK4)
CHEK2	CHK2 checkpoint homolog (S. pombe)
CORO1B	coronin, actin binding protein, 1B
CREB1	cAMP responsive element binding protein 1
CREBBP	CREB binding protein (Rubinstein-Taybi syndrome)
CSE1L	CSE1 chromosome segregation 1-like (yeast)
CSNK2A1	casein kinase 2, alpha 1 polypeptide
CTNNB1	catenin (cadherin-associated protein), beta 1, 88kDa
CUL7	cullin 7
DUSP1	dual specificity phosphatase 1
DUSP16	dual specificity phosphatase 16
DYRK2	dual-specificity tyrosine-(Y)-phosphorylation regulated kinase 2
E2F1	E2F transcription factor 1
E4F1	E4F transcription factor 1
EEF1A1	eukaryotic translation elongation factor 1 alpha 1
EIF4EBP1	eukaryotic translation initiation factor 4E binding protein 1
ESR1	estrogen receptor 1
ESR2	estrogen receptor 2 (ER beta)
FAS	Fas (TNF receptor superfamily, member 6)
FCGR1A	Fc fragment of IgG, high affinity Ia, receptor (CD64)
FHL2	four and a half LIM domains 2
FLT1	fms-related tyrosine kinase 1 (vasc. endoth. growth factor/vasc. permeability factor rec.)
FNIP1	folliculin interacting protein 1
FOS	v-fos FBJ murine osteosarcoma viral oncogene homolog
FOXO1A	forkhead box O1
FRAP1	FK506 binding protein 12-rapamycin associated protein 1
GADD45A	growth arrest and DNA-damage-inducible, alpha
GTF2H2	general transcription factor IIH, polypeptide 2, 44kDa
GTSE1	G-2 and S-phase expressed 1
HIPK2	homeodomain interacting protein kinase 2

Table 3.1.1 (continued)

IGF1	insulin-like growth factor 1 (somatomedin C)
IGFBP3	insulin-like growth factor binding protein 3
JMY	junction-mediating and regulatory protein
KDR	kinase insert domain receptor (a type III receptor tyrosine kinase)
LRDD	leucine-rich repeats and death domain containing
MAP2K4	mitogen-activated protein kinase kinase 4
MAP3K5	mitogen-activated protein kinase kinase kinase 5
MAP4	microtubule-associated protein 4
MAPK1	mitogen-activated protein kinase 1
MAPK8	mitogen-activated protein kinase 8
MAPK9	mitogen-activated protein kinase 9
MAPT	microtubule-associated protein tau
MDM2	MDM2 p53 binding protein homolog (mouse)
MDM4	Mdm4 p53 binding protein homolog (mouse)
MMP2	matrix metallopeptidase 2 (gelatinase A, 72kDa gelatinase, 72kDa type IV collagenase)
MTA2	metastasis associated 1 family, member 2
MYST1	MYST histone acetyltransferase 1
NFKB1	nuclear factor of kappa light polypeptide gene enhancer in B-cells 1
NUMB	numb homolog (Drosophila)
p53AIP1	p53-regulated apoptosis-inducing protein 1
PARC	p53-associated parkin-like cytoplasmic protein
PARP1	poly (ADP-ribose) polymerase 1
PCAF	K(lysine) acetyltransferase 2B
PDK1	pyruvate dehydrogenase kinase, isozyme 1
PIAS1	protein inhibitor of activated STAT, 1
PIAS4	protein inhibitor of activated STAT, 4
PIK3CA	phosphoinositide-3-kinase, catalytic, alpha polypeptide
PIK3R1	phosphoinositide-3-kinase, regulatory subunit 1 (alpha)
PIK3R2	phosphoinositide-3-kinase, regulatory subunit 2 (beta)
PIN1	peptidylprolyl cis/trans isomerase, NIMA-interacting 1
PMAIP1	phorbol-12-myristate-13-acetate-induced protein 1
POU4F1	POU class 4 homeobox 1
POU4F2	POU class 4 homeobox 2
PPM1D	protein phosphatase 1D magnesium-dependent, delta isoform
PPP1R13B	protein phosphatase 1, regulatory (inhibitor) subunit 13B
PPP1R13L	protein phosphatase 1, regulatory (inhibitor) subunit 13 like
PPP2CA	protein phosphatase 2 (formerly 2A), catalytic subunit, alpha isoform
PPP2CB	protein phosphatase 2 (formerly 2A), catalytic subunit, beta isoform
PPP2R1A	protein phosphatase 2 (formerly 2A), regulatory subunit A , alpha isoform
PPP2R1B	protein phosphatase 2 (formerly 2A), regulatory subunit A, beta isoform
PPP2R2A	protein phosphatase 2 (formerly 2A), regulatory subunit B, alpha isoform
PPP2R2B	protein phosphatase 2 (formerly 2A), regulatory subunit B, beta isoform
PPP2R2C	protein phosphatase 2 (formerly 2A), regulatory subunit B, gamma isoform
PPP2R3A	protein phosphatase 2 (formerly 2A), regulatory subunit B'', alpha
PPP2R3B	protein phosphatase 2 (formerly 2A), regulatory subunit B'', beta
PPP2R4	protein phosphatase 2A activator, regulatory subunit 4
PPP2R5A	protein phosphatase 2, regulatory subunit B', alpha isoform
PPP2R5B	protein phosphatase 2, regulatory subunit B', beta isoform
PPP2R5C	protein phosphatase 2, regulatory subunit B', gamma isoform
PPP2R5D	protein phosphatase 2, regulatory subunit B', delta isoform

Table 3.1.1 (continued)

PPP2R5E	protein phosphatase 2, regulatory subunit B', epsilon isoform
PRKAA1	protein kinase, AMP-activated, alpha 1 catalytic subunit
PRKAB1	protein kinase, AMP-activated, beta 1 non-catalytic subunit
PRKAB2	protein kinase, AMP-activated, beta 2 non-catalytic subunit
PRKAG2	protein kinase, AMP-activated, gamma 2 non-catalytic subunit
PSRC1	proline/serine-rich coiled-coil 1
PTEN	phosphatase and tensin homolog
RAB13	RAB13, member RAS oncogene family
RB1	retinoblastoma 1
RBBP4	retinoblastoma binding protein 4
RCHY1	ring finger and CHY zinc finger domain containing 1
RFWD2	ring finger and WD repeat domain 2
RHEB	Ras homolog enriched in brain
RPRM	reprimo, TP53 dependent G2 arrest mediator candidate
RPS6KB1	ribosomal protein S6 kinase, 70kDa, polypeptide 1
RRM2	ribonucleotide reductase M2 polypeptide
SERPINB2	serpin peptidase inhibitor, clade B (ovalbumin), member 2
SERPINB5	serpin peptidase inhibitor, clade B (ovalbumin), member 5
SFN	stratifin
SIAH1	seven in absentia homolog 1 (Drosophila)
SKP2	S-phase kinase-associated protein 2 (p45)
STEAP3	STEAP family member 3
STK11	serine/threonine kinase 11
STMN1	stathmin 1/oncoprotein 18
STRAP	serine/threonine kinase receptor associated protein
SUMO1	SMT3 suppressor of mif two 3 homolog 1 (S. cerevisiae)
TDG	thymine-DNA glycosylase
THBS1	thrombospondin 1
TNFRSF10B	tumor necrosis factor receptor superfamily, member 10b
TP53	tumor protein p53
TP53BP1	tumor protein p53 binding protein 1
TP53BP2	tumor protein p53 binding protein, 2
TP63/TP73L	tumor protein p63
TP73	tumor protein p73
TSC1	tuberous sclerosis 1
TSC2	tuberous sclerosis 2
USP7	ubiquitin specific peptidase 7 (herpes virus-associated)
VEGFA	vascular endothelial growth factor A
WNT1	wingless-type MMTV integration site family, member 1
WWP1	WW domain containing E3 ubiquitin protein ligase 1
XPA	xeroderma pigmentosum, complementation group A
XPC	xeroderma pigmentosum, complementation group C
XPO1	exportin 1 (CRM1 homolog, yeast)
YWHAQ	tyrosine 3-monooxygenase/tryptophan 5-monooxyg. activation protein, theta polypeptide
ZNF385A	zinc finger protein 385A

Table 3.1.2. Twenty-nine HapMap PPP2R5E SNPs with minor allele frequencies greater than 10% have extreme iHS scores[1]. *ε-SNP2

Region	Contig Position	SNP	Alleles	Derived Allele	Derived Allele Frequency	iHS
intron 2	44978228	rs1255641	C/T	C	0.885	2.55
intron 2	44965697	rs1255627	A/G	A	0.839	2.80
intron 2	44949747	rs17101239	C/G	C	0.780	2.74
intron 2	44948642	rs10145422	C/T	T	0.828	2.61
intron 2	44945051	rs1951213	C/G	G	0.828	2.61
intron 2	44944616	rs11158492	A/G	G	0.828	2.61
intron 2	44942312	rs7153764	C/T	C	0.828	2.71
intron 2	44941827	rs1957045	A/G	G	0.828	2.71
intron 2	44941305	rs4391977	A/G	G	0.828	2.89
intron 2	44936970	rs8018107	A/T	T	0.828	2.79
intron 2	44934705	rs1033745	C/T	C	0.839	2.57
intron 2	44933964	rs1033744	C/T	C	0.833	2.78
intron 2	44931618	rs10141179	C/T	C	0.828	2.75
intron 2	44926433	rs7143816	C/T	T	0.828	2.75
intron 2	44923835	rs12323479	A/G	G	0.828	2.76
intron 2	44922047	rs10151527	A/C	C	0.722	-3.00
intron 2	44922013	rs4902226	C/T	T	0.822	2.88
intron 3	44916716	rs10152067	A/G	A	0.828	2.60
intron 3	44914073	rs6573518	C/T	T	0.828	2.61
intron 3	44910858	rs11158491*	C/T	C	0.722	-2.87
intron 3	44909091	rs7146069	A/G	G	0.806	2.84
intron 3	44907997	rs8003647	C/T	T	0.828	2.65
intron 3	44907693	rs8018307	C/G	C	0.828	2.65
intron 3	44901047	rs10140679	C/T	C	0.744	-2.65
intron 4	44888075	rs10145042	C/T	T	0.911	2.68
intron 5	44880199	rs7141605	C/G	C	0.828	2.94
intron 5	44879791	rs7140704	A/G	A	0.844	2.76
intron 5	44876923	rs11158490	G/T	T	0.833	2.85
intron 11	44853005	rs7146445	A/C	A	0.906	2.53

The subsequent analyses focus on the three SNPs wherein the derived alleles associate with unusually long haplotypes as they might denote a functional nonancestral allele that has undergone a recent selective sweep. The alleles of all three loci are highly linked to each other, to such an extent that these associations are nonrandom and, therefore, in very high LD. Specifically, rs10151527 in intron 2 (ε-SNP1) is in either complete or almost complete LD with two SNPs in intron 3, rs11158491 (ε-SNP2; D′=1;

[1] Alexei Vazquez

r^2=1) and rs10140679 (ε-SNP3; D′=1; r^2=0.87). Indeed, the derived alleles for all three SNPs are at very high frequencies in the central European populations (>72%), and the low diversity of the derived haplotypes are readily apparent. As seen in Figure 3.1.1, when we analyze the marginal haplotypes of the 15 SNPs centered on ε-SNP2 (rs11158491), it is clear that the high-frequency (73%) derived allele (C) shows significantly lower haplotype diversity than the lower-frequency (27%) ancestral allele (T), and has a dominant haplotype in 69% of the population. This noted high-frequency, low-diversity haplotype is similar to the haplotype containing the functional G-allele of MDM2 SNP309 (Atwal, Bond et al. 2007). We reasoned that the PPP2R5E haplotype, denoted by these SNPs, might also harbor one or more functional alleles and, therefore, also show allelic differences in human cancer populations.

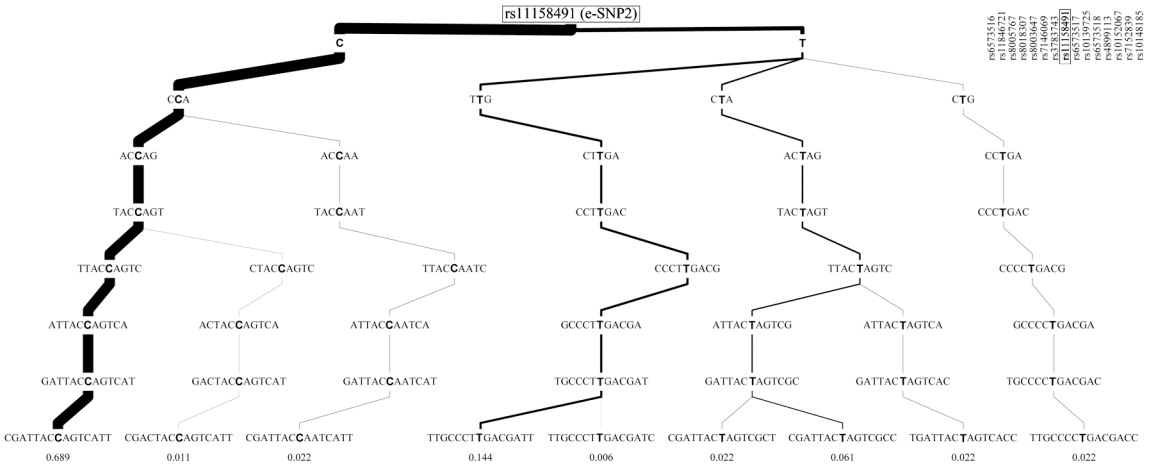


Figure 3.1.1. A depiction of the haplotype structure of the PPP2R5E gene in Central Europeans centered on ε-SNP2 (rs11158491). The C-allele is the derived allele of PPP2R5E ε-SNP2 and the T allele is the ancestral allele. The numbers at the bottom of each haplotype denote the inferred frequencies. In the upper right corner, the order of the SNPs is reported using the annotation of dbSNP[1].

[1] Alexei Vazquez

3.1.2.1. A selected SNP in the PPP2R5E gene associates with altered sarcoma risk and survival

The first noted clinical association of the G-allele of MDM2 SNP309 was its association with an earlier tumor onset in soft-tissue sarcoma patients, first in p53 mutation carriers and subsequently in sporadic cases (Bond, Hu et al. 2004). As mentioned in the Introduction (Chapter 1.4.2), similar effects have been subsequently observed in a variety of other tumor types (Bond and Levine 2007). To test the hypothesis that the PPP2R5E haplotype might also harbor one or more functional SNPs, which show allelic differences in human cancer populations, we genotyped ε-SNP2 in the same cohort of sporadic STS patients in which the affect of age-dependent incidence of the MDM2 SNP309 locus was first described (Bond, Hirshfield et al. 2006; Bond and Levine 2007).

Similar to MDM2 SNP309, when the sarcoma patients were separated based on the PPP2R5E ε-SNP2 genotype, patients homozygous for the T-allele were found to be diagnosed significantly earlier in life than individuals C/T or C/C in genotype. Specifically, sarcoma patients with a C/C and C/T genotype were diagnosed with a mean age of 58.2 and 57.6 years, respectively, whereas patients homozygous for the T-allele showed an 11.8-year earlier age of tumor onset at 46.4 years ($p=0.001$, permutation test; Table 3.1.3). Interestingly, a further similarity with the observations made for MDM2 SNP309 (Bond, Hirshfield et al. 2006; Bond, Menin et al. 2006; Bond and Levine 2007; Atwal, Rabadan et al. 2008) was that these allelic differences show a strong gender-dependence. For ε-SNP2, the greatest allelic difference in the age of onset was found in

the male patients. Males T/T in genotype showed a 19.2-year earlier tumor onset when compared with males homozygous for the C-allele (p=0.0002, permutation test; Table 3.1.3). A similar, albeit weaker, 5.7-year difference was observed in the female patients (Table 3.1.3). However, these differences failed to reach statistical significance.

Table 3.1.3. Average (median) age of tumor diagnosis for PPP2R5E ε-SNP2 genotypes in years.

	n	CC	CT	TT	P-value*
Total	130	58.2 (59)	57.6 (62)	46.4 (44)	0.0010
Males	57	59.1 (62)	51.7 (57.5)	39.9 (41.5)	0.0002
Females	73	57.3 (57.5)	61.2 (65)	51.6 (53.5)	>0.05

*Permutation Test

To expand our analysis of the allelic differences of ε-SNP2 in the risk of developing soft-tissue sarcomas, we genotyped ε-SNP2 in 497 blood donors of the same ethnic background. Among them, 38 (T/T, 7.6 %) were homozygous for the ε-SNP2 T-allele, 266 (C/C, 53.5 %) were homozygous for the C-allele, and 193 (T/C, 38.9 %) were heterozygous (Table 3.1.4). These ε-SNP2 genotype frequencies differed significantly from those found in the 130 sarcoma patients. Interestingly, an enrichment of individuals homozygous for the T-allele (13.8%) and a decrease in the C/C genotype frequency (50.0%) were noted in the sarcoma patient group, suggesting that T/T individuals harbor a higher risk for sarcoma development (OR=2.3; p=0.025, Bayesian estimate; Table 3.1.4)

Table 3.1.4. Summary of case-control analyzes of PPP2R5E ε-SNP2.

	n	C/C*	C/T*	T/T*	p-value**	Mean Odds Ratio*** T/T: C/C	95% CI***
Controls							
Total	497	266 (53.5)	193 (38.9)	38 (7.6)	-	-	-
Cases							
Total	130	65 (50.0)	47 (36.2)	18 (13.8)	0.0250	2.3	1.05 - 3.67
< 51 yrs	43	19 (44.2)	12 (27.9)	12 (27.9)	0.0010	5.3	1.82 - 9.58
and < 66 yrs	43	23 (53.5)	17 (39.5)	3 (7.0)	>0.05		
≥ 66 yrs	44	23 (52.3)	18 (40.1)	3 (6.8)	>0.05		
Males	57	31 (54.5)	18 (31.5)	8 (14.0)	>0.05		
< 51 yrs	21	8 (38.1)	6 (28.6)	7 (33.3)	0.004	8.1	1.57 - 17.23
and < 66 yrs	20	12 (60.0)	8 (40.0)	0 (0)	>0.05		
≥ 66 yrs	16	11 (68.7)	4 (25.0)	1 (6.3)	>0.05		
Females	73	34 (46.6)	29 (39.7)	10 (13.7)	>0.05		
< 51 yrs	22	11 (50.0)	6 (27.3)	5 (22.7)	0.048	6.1	0.87 - 13.65
and < 66 yrs	23	11 (47.8)	9 (39.2)	3 (13.0)	>0.05		
≥ 66 yrs	28	12 (42.9)	14 (50.0)	2 (7.1)	>0.05		

* n (%)
** Probability to observe (n) TTs in (n) cases given the (n) TTs in (n) controls. Cases and Controls were age and gender matched.
***Bayesian estimate

The above-described age-of-onset analysis suggests that significant enrichments of particular genotypes in the patient populations, compared with controls, could also be gender- and/or age-dependent. Therefore, incorporating these parameters could identify individuals with an even higher risk of sarcoma development. To do this, we divided either all sarcoma patients, or male patients and female patients, into three groups of almost equal size according to their age at diagnosis: the youngest, intermediately aged, and oldest. Interestingly, upon comparison of the genotype frequencies, the youngest patients in all three groupings (total, males and females) showed a significant enrichment of the T/T genotype when compared with the controls, and a subsequent decrease in the frequency of the C/C genotype (Table 3.1.4). Specifically, individuals homozygous for the T-allele made up 27.9% of the youngest patients as opposed to 7% of intermediately

aged, 6.8% of oldest cases, and 7.6% of controls (OR=5.3; p=0.0010; Table 3.1.4). Furthermore, in agreement with the previous results, these observations were particularly pronounced in the youngest male patients, whereas men T/T in genotype made up 33.3% of the male cases diagnosed at youngest age, as opposed to 6.3% of oldest male cases, and 7.6 % of blood donors (OR=8.1; P=0.0040; Table 3.1.4). In contrast, men C/C in genotype made up only 38.1% of the male cases diagnosed at youngest age, as opposed to 60% of intermediately aged, 68.7% of oldest male cases, and 53.5% of blood donors. As with the age-of-onset analysis, similar, weaker trends were observed in the female patients, whereas the T/T genotype made up 22.7% of youngest females (OR=6.1; P=0.048; Table 3.1.4) and only 7.1% of oldest women.

Together, the above-described significant allelic differences in age-dependent onset and risk of STS development lend strong support to the hypothesis that the PPP2R5E haplotypes, denoted by PPP2R5E ε-SNP2, might also harbor one or more functional, clinically relevant SNP(s). To further test the perceived clinical relevance of ε-SNP2, we assessed the potential impact of this locus on the overall survival of the sarcoma patients. To begin with, we did a Kaplan-Meier survival analysis that revealed that sarcoma patients C/C in genotype associated with the worst overall survival with a mean survival time of 48.4 months, followed by patients heterozygous for the C-allele (mean survival time of 69.9 months) and subsequently individuals with a T/T genotype (mean survival time of 142.7 months; p=0.034, log rank test; Figure 3.1.2A). To exclude potential biases from other independent prognostic factors, we carried out a multivariate Cox's regression survival analysis adjusted to tumor stage, type of tumor resection and gender (Molife, Lorigan et al. 2001; Lahat, Tuvin et al. 2008). This analysis further

confirmed the initial results, showing a 2.68-fold increased risk for tumor-related death for patients C/C in genotype when compared with individuals homozygous for the T-allele (p=0.045; Figure 3.1.2).

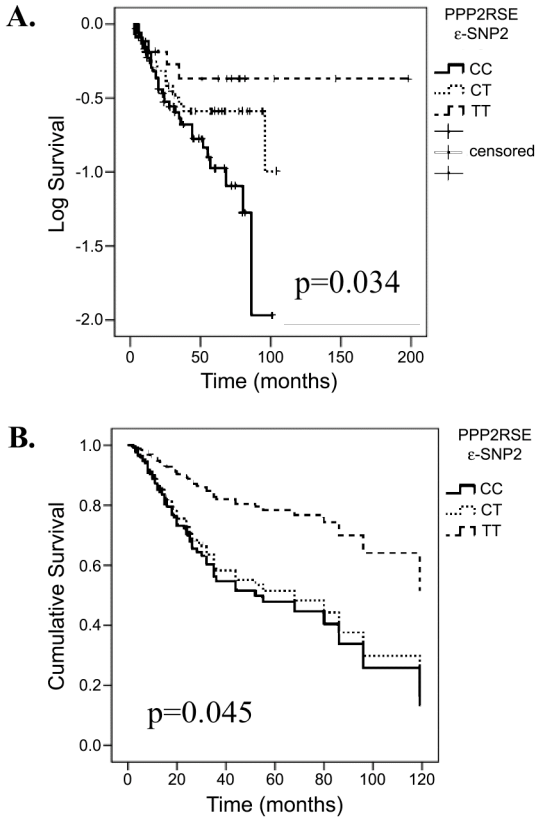


Figure 3.1.2. Survival analysis of PPP2R5E ε-SNP2 shows an association of the derived C-allele with an increased risk for tumor-related death. The survival curves of soft-tissue sarcoma patients for each genotype are plotted against the survival time in months. Patients who were alive at the time of follow-up were included as censored data into the analyses. The results of the Kaplan-Meier analysis (A) are further supported by the Cox multivariate regression survival analysis (B) that was adjusted to tumor stage, type of tumor resection and gender. In the Cox analysis, patients with a C/C genotype (black line) showed a 2.68-fold increased risk for tumor-related death when compared with patients homozygous for the T-allele (black dashed line).

3.1.3. Discussion

In this report, we set out to search for genetic variants in 142 genes known to be important in mediating and regulating the p53 stress response that show signs of natural selection and differences in sarcoma risk and survival. By utilizing a map of recent natural selection (Haplotter), we found that only one of these 142 genes resides in a genomic region that shows highly significant evidence of selection. This gene (PPP2R5E) encodes a regulatory subunit of PP2A. The strongest signs of natural selection that associated with derived (nonancestral) alleles were from three SNPs that were in high linkage disequilibrium with each other. To assess potential differences in sarcoma risk and survival, one of these linked SNPs (PPP2R5E ε-SNP2) was genotyped in a cohort of patients diagnosed with soft-tissue sarcoma and a control group of the same ethnicity. Interestingly, the subsequent analysis revealed that, in STS patients, PPP2R5E ε-SNP2 associates strongly with both altered risk and survival. Individuals homozygous for the ancestral allele (T/T) associate with significant allelic differences in the onset and risk of STS development in a gender-dependent manner. Specifically, male T/T carriers are diagnosed 19 years earlier than male C/C carriers and have a significantly increased risk of developing early onset disease (OR=8.1; Table 3.1.4). However, once diagnosed with soft-tissue sarcoma, C/C carriers associate with a significant and increased risk of tumor-related death (relative risk, 2.68; Fig. 3.1.2).

The genomic region noted to be under evolutionary selection pressure also included two other, less-characterized genes, WDR89 and SGPP1. Neither gene has been shown to play a role in the induction or regulation of malignant transformation. In

contrast, mounting evidence of the tumor suppressing activity of the PP2A holoenzyme can be found in the literature (Chen, Arroyo et al. 2005; Mumby 2007; Westermarck and Hahn 2008). An important mechanism of intracellular signaling is the regulated, reversible phosphorylation of proteins, and deregulation of these events can play a central role in the malignant transformation of the cell. Recent work has provided compelling evidence for the role of specific protein phosphatases, particularly of PP2A, in carcinogenesis and tumor progression (Mumby 2007; Westermarck and Hahn 2008). PP2A phosphatase activity has been shown to interact directly with the p53 pathway, causing the dephosphorylation of key residues of p53 and MDM2, resulting in the regulation of p53 activity levels in cells (Kimura and Nojima 2002; Okamoto, Li et al. 2002; Dohoney, Guillerm et al. 2004; Moule, Collins et al. 2004; Li, Cai et al. 2007). For example, the PP2A-mediated dephosphorylation of MDM2 has been linked to the ability of MDM2 to degrade p53 (Okamoto, Li et al. 2002). It will be interesting to determine whether the varying levels of MDM2 protein produced by the different alleles of MDM2 SNP309 can modify the significant allelic differences of the PPP2R5E variant in cancer risk and survival.

The genes that encode the subunits of the PP2A holoenzyme were included in our study due to the above-described role that PP2A plays in regulating p53 signaling (Kimura and Nojima 2002; Okamoto, Li et al. 2002; Dohoney, Guillerm et al. 2004; Moule, Collins et al. 2004; Li, Cai et al. 2007). However, PP2A can affect other, at least partially, p53-independent intracellular signaling pathways that affect human cancer (Westermarck and Hahn 2008). For example, PP2A can antagonize Ras signaling by dephosphorylating c-Myc and RalA and by negatively regulating the PI3K/Akt signaling

pathway (Mumby 2007). Indeed, it will be of future interest to explore whether the significant allelic differences in haplotype structure, cancer risk, and cancer survival noted in this report are due to allelic differences in p53 signaling or other signaling pathways affected by PP2A activity. Further evidence of the importance of PP2A activity in suppressing cellular transformation lies in the wide range of mechanisms that transformed cells have evolved to inhibit its activity. For example, inhibition of PP2A activity has been shown to be mediated by the small tumor antigen of DNA tumor viruses (Arroyo and Hahn 2005) by up-regulation of the c-Myc–specific inhibitor CIP2A (Junttila, Puustinen et al. 2007), by BCR/ABL via SET up-regulation (Neviani, Santhanam et al. 2005), through biallelic mutational inactivation of the Aβ subunit (Sablina, Chen et al. 2007), or by decreased expression of the Aα subunit (Chen, Arroyo et al. 2005).

PP2A is a ubiquitously expressed heterotrimeric protein that accounts for a large fraction of phosphatase activity in eukaryotic cells (Mumby 2007). The various forms of PP2A contain an active core dimer, made up of the catalytic (C-) subunit and a structural scaffold (A-) subunit. The scaffold subunit mediates interaction of the core dimer with a great variety of regulatory (B-) subunits that determine the localization, specificity, and functions of individual isoforms (Westermarck and Hahn 2008). Four families of B subunits have been described: B/B55/PR55/PPP2R2, B′/B56/PR61/PPP2R5, B″/PR72/PPP2R3, and B‴/PR93/SG2NA/PR110/Striatin (Mumby 2007). Furthermore, five different PPP2R5 genes have been identified: PPP2R5A (α-), PPP2R5B (β-), PPP2R5C (γ-), PPP2R5D (δ-) and PPP2R5E (ε-isoform; (Martens, Stevens et al. 2004)). To date, little is known about the specific regulatory role of PPP2R5E. However, one

recent report provided evidence that PPP2R5E can have both p53-dependent and – independent anti- and pro-apoptotic functions during development of xenopus embryos. (Jin, Wallace et al. 2010).

The observations reported here remain to be validated in other patient cohorts, and importantly, the regulatory changes associated with PPP2R5E ε-SNP2 are yet to be determined. As we will describe in more detail in Chapter 3.4, the SNPs that associated with the strongest signs of natural selection with derived (nonancestral) alleles are in high linkage disequilibrium with each other and reside in either intron 2 or intron 3 (Table 3.1.2). To our knowledge, there have been no published attempts to characterize the regulatory regions of the PPP2R5E gene. However, we will aim to begin to further elucidate the potential cellular mechanisms behind the significant allelic differences in haplotype structure, cancer risk, and cancer survival in Chapter 3.4. Therein, we will present some evidence that those differences might be due to allele-specific regulation of PPP2R5E transcript expression in cells. In summary, the data reported in this Chapter suggest that PPP2R5E plays a significant regulatory role in the development and progression of soft-tissue sarcomas and harbors genetic variants that can affect human cancer and may be under evolutionary selection pressure.

3.2. Chemosensitivity profiles identify polymorphisms in the p53 network genes 14-3-3τ and CD44 that affect sarcoma incidence and survival

3.2.1. Introduction

3.2.2. Results

3.2.2.1. SNPs in the 14-3-3τ and CD44 genes associate with allelic differences in cellular drug responses

3.2.2.2. SNP alleles with reduced chemosensitivities associate with worse sarcoma survival and earlier sarcoma incidence

3.2.3. Discussion

3.2.1. Introduction

Genes that constitute the p53 network do not only affect cancer susceptibility, but are also critical in the regulation of the cellular response to treatment with DNA-damaging therapeutic agents. Specifically, radio- and chemotherapeutic agents lead to the activation of the of the p53 stress response that can result in cancer cell-clearance through activation of apoptotic or senescence pathways (Chapter 1.1.2; (Lowe and Lin 2000; Zhou and Elledge 2000; Johnstone, Ruefli et al. 2002; Bolderson, Richard et al. 2009). Indeed, mutations in the p53 gene and other genes from this important tumor suppressor network significantly affect cellular chemosensitivities in experimental models and are associated with multidrug resistance in human cancers (Chapter 1.1.2; (Reed 1999; Wallace-Brodeur and Lowe 1999; Schmitt, Rosenthal et al. 2000; Zhang, Yu et al. 2000; Wei, Zong et al. 2001; Johnstone, Ruefli et al. 2002).

A body of evidence is emerging in the literature that suggests that the inherited genetics of the p53 pathway could be used to further define patient populations in their abilities to respond to stress, suppress tumor formation, and induce p53 activity in response to DNA-damaging therapies (Bond and Levine 2007; Vazquez, Bond et al. 2008). Specifically, the different alleles of p53 codon72 encode either a proline or arginine residue, which reside in a proline-rich region of p53 that has been shown to be important in mediating the apoptotic response (Chapter 1.4.1; (Sakamuro, Sabbatini et al. 1997; Vazquez, Bond et al. 2008; Zhu, Dolle et al. 2010; Frank, Leu et al. 2011). Furthermore, in tumor-derived cell lines, the elevated levels of MDM2 associated with the G-allele of MDM2 SNP309 have been shown in multiple studies to result in the

attenuation of the p53 apoptotic response after exposure to multiple types of chemotherapeutics compared with cells containing the T-allele (Chapter 1.4.2; (Bond, Hu et al. 2004; Arva, Gopen et al. 2005; Nayak, Yang et al. 2007). These observations for both p53 codon72 and MDM2 SNP309 correlate well with studies of many types of cancers for which allelic differences have been noted for these loci in the onset of and risk for cancer, as well as in the response to therapy and survival (Bond and Levine 2007; Vazquez, Bond et al. 2008; Whibley, Pharoah et al. 2009).

The above-mentioned studies suggest that this important stress response pathway could harbor more functional inherited genetic variants, the study of which could help further define patient populations in their abilities to respond to stress, suppress tumor formation, and respond to DNA-damaging therapies. Recently, Vazquez et al proposed a methodology to uncover candidate functional SNPs using data derived from the NCI60 human tumor cell line anticancer drug screen (Vazquez, Bond et al. 2008). NCI60 is a panel of 59 cancer cell lines from different tissues of origin (Chapter 2.1.3.2; (Shoemaker 2006). Much is known of their genetics: the mutational status of more than 20 important cancer-related genes and the genotypes of 100,000 SNPs (Garraway, Widlund et al. 2005; Ikediobi, Davies et al. 2006). A great deal is also known of their response to therapies: both the growth response of these cell lines to more than 100 standard chemotherapeutic agents as well as to more than 40,000 publicly available potential anticancer agents (Shoemaker 2006). The growth response was measured 48 hours after treatment at a five-log concentration range for each compound. The negative $\log_{10}$ (GI_{50}) of the concentration required to inhibit the growth of the given cell line by 50% relative to the vehicle-treated cells is calculated and reported (Shoemaker 2006). It is the most

commonly used and most sensitive summary measure of a dose-response curve in functional antagonist assays (Sebaugh 2010). A previous study showed that the cells with mutations in the p53 gene are less sensitive to many standard chemotherapeutic agents that induce DNA damage, suggesting that the cellular response to these agents is p53-dependent (O'Connor, Jackman et al. 1997). Vazquez et al developed an analytic framework that they implemented to mine the publicly available genotype, mutational, and cellular drug response data to identify SNPs in the p53 stress response that merited future study (Vazquez, Bond et al. 2008). In this Chapter, we build upon, expand, and use this methodology and identify genetic variants in the CD44 and YWHAQ (14-3-3τ) genes that significantly affect both the cellular responses to many chemotherapeutics, as well as human cancer incidence and survival. The work presented herein is a result of the collaboration with Alexei Vazquez from The Institute for Advanced Study in Princeton, New Jersey, USA and has been published during the continued work on my Thesis (Vazquez, Grochola et al. 2010).

3.2.2. Results

3.2.2.1. SNPs in the 14-3-3τ and CD44 genes associate with allelic differences in cellular drug responses

The genotypes for more than 100,000 SNPs are publicly available for the NCI60 cell line panel and were determined using the Affymetrix 125K genotyping platform (http://dtp.nci.nih.gov/). Two hundred sixty-four genotyped SNPs were found to reside in the 142 genes known to be important in the p53 stress response (Table 3.1.1, Chapter

3.1). We set out to explore possible allelic differences of SNPs residing in these genes in their growth responses to 132 standard chemotherapeutic agents. To do this, three statistical tests were performed (Figure 3.2.1A). First, a univariate test was undertaken for each drug and SNP pair and allelic differences were sought. Specifically, the average log GI50 [$X = -\log_{10}(GI_{50})$] for cells for each of the three genotypes of a given locus (AA, Aa, and aa) were calculated for cells either wild-type or mutant for p53 (Materials and Methods, Table 2.8; O'Connor, Jackman et al. 1997). Subsequently, the probability (p-value) was computed that just by chance the difference for the following groupings either was equal to or larger than the actual measurement: (a) Xa-X_{AA} or (b) X_{aa}-X_A, or (c) X_{AA}-X_a, or (d) X_A-X_{aa}, or (e) [X_{aa}-X_{aA} and X_{aA}-X_{AA}], and (f) [X_{AA}-X_{aA} and X_{aA}-X_{aa}]. These probabilities were estimated using a permutation test (106 permutations) that preserved the allele or genotype group sizes but permuted the samples among the groups. Results $p<0.05$ were considered significant. Four different combinatorial hits were defined for each SNP-drug pair: 1, significant in all cells (SA); 2, significant in cells with a wild-type p53 gene (SWT); 3, significant in cells with mutant p53 (SMT); and 4, significant in cells with wild-type p53 but not significant or significant in the opposite allele orientation in cells with a mutant p53 (SWT-NRMT; Fig. 3.2.1B).

Second, a multiple hypothesis test was performed for the noted observations (Figure 3.2.1A). This test took advantage of the fact that 132 well-characterized compounds were tested against the NCI60 cell panel, which provided a set of independent measurements. The test incorporated all 264 p53 stress response SNPs and 132 agents and used a Fisher's exact test to compute the statistical significance of observing h univariate hits for a SNP (assuming one of the possibilities 1–4 listed above)

on a total of D drugs, given that overall H significant hits are observed after testing S reference SNPs on the D drugs. We chose all 109,687 genotyped SNPs as a reference set.

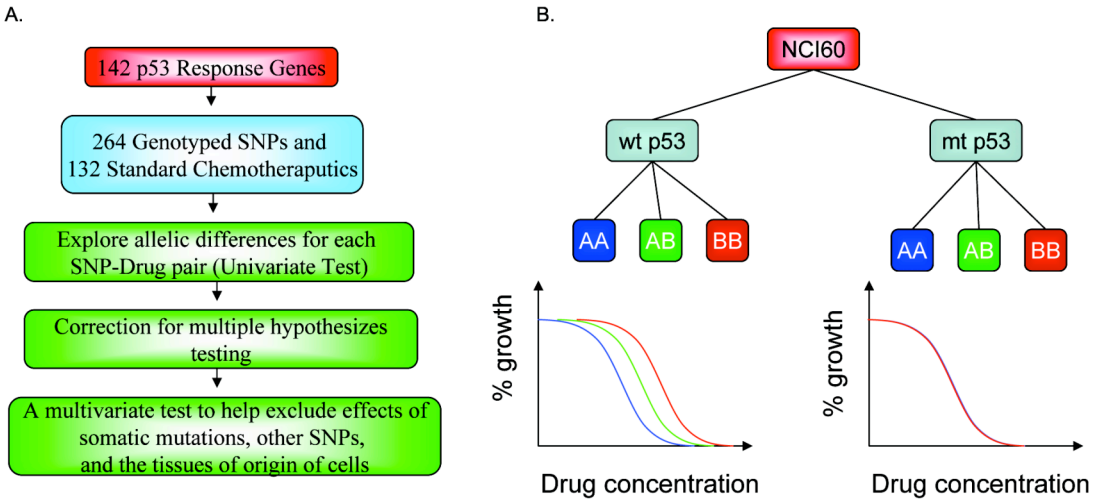


Figure 3.2.1. Schematic representation of the chemosensitivity screen in the NCI60 cell panel. (A) Flowchart depicting the statistical approaches designed to identify SNPs that associate with allelic differences in chemosensitivity in the NCI60 human tumor anticancer drug screen. (B) A graph representing the stratification of the NCI60 cell lines according to p53 mutational status and SNP genotype.

The third statistical test served to help eliminate the possibility that a given allelic difference in the growth response to an agent could be due to the tumor type of the cells, other somatic mutations or other SNPs, or a combination of these variables (Figure 3.2.1A). To do this, a two-step multivariate test was performed. Using a log-linear regression model, the contribution of each attribute [tissue of origin, somatic mutational status of 19 cancer-related genes (APC, BRAF, BRCA1, BRCA2, CDH1, CDKN2A, CTNNB1, EGFR, ERBB2, FLT3, HRAS, KRAS, NRAS, MADH4, PDGFRA, PIK3CA, PTEN, RB1, SKT11, VHL), and the SNP genotypes for the 264 p53 stress response SNPs] to the log GI_{50} over the set of 132 drugs was computed. In the second step, identical computations were performed on a second set of 264 SNPs that were chosen

randomly from the 109,687 genotyped SNPs. These calculations served as a reference set that allowed for the estimation of the statistical significance of the effects observed for a given SNP within the p53 stress response, specifically by defining the probability (p-value) that a SNP in the reference set has an equal or larger contribution to the observed effect. The SNPs within p53 stress response with at least two genotypes demonstrating effects with p-values lower than 5% in cells with wild-type p53 were deemed significant.

Only 8 of the 264 SNPs, located in the YWHAQ (rs6734469), PPP2R2B (rs319227 and rs319217), CCNG1 (rs2069347), CD44 (rs187115), PIAS1 (rs1027154), KDR (rs2168945) and CSEIL (rs2426127) genes, showed significant allelic differences in cellular growth responses to the standard chemotherapeutic agents according to the SWT-NRMT univariate analysis, the associated multiple hypothesis test, and the multivariate test in the cells wild-type for p53 (Vazquez, Grochola et al. 2010). To complete and validate the genotyping of the eight candidate SNPs, we obtained the genomic DNAs from all 59 cell lines and re-genotyped each SNP using accurate allelic discrimination assays (Applied Biosystems). The publicly available genotypes derived from the Affymetrix 125K genotyping platform had, on average, nine missing genotypes and, compared with the re-genotyping, six missed calls (Vazquez, Grochola et al. 2010). Therefore, we repeated the above-described genotype-response analysis using the new genotypes. All but one (rs2426127 in CSEIL) of the eight SNPs showed significant allelic differences in cellular growth responses to the standard chemotherapeutic agents in the cells that were wild-type for p53 (Table 3.2.1). Specifically, on average, the different genotypes of the seven significant SNPs associate with a different growth response for 79 (range 57–108; Table 3.2.1) of the 132 agents tested, whereby cells with the homozygote

genotypes differed on average 4.9-fold (range 3.4- to 7.7-fold; Table 3.2.1) in their respective drug sensitivities in p53 wild-type cells.

The strongest effects in cells wild-type for p53 were seen for the SNP in the YWHAQ (14-3-3τ) gene (rs6734469; Table 3.2.1). Specifically, the results of the univariate analysis suggest that cells with the different genotypes of the YWHAQ SNP significantly differ in their growth response to 108 of the 132 agents tested in wild-type p53 cells ($p=5.6\times10^{-47}$; Table 3.2.1), but only 3 of the 132 agents in mutant p53 cells. Significant allelic differences were observed with 35 alkylating agents, 20 and 15 topoisomerase I and II inhibitors, 14 and 15 RNA/DNA and DNA antimetabolites, and, to a lesser extent, with 10 antimitotic agents (Figure 3.2.2A and B). Wild-type p53 cells with the homozygote genotypes for YWHAQ SNP differed on average 5.4-fold in their respective drug sensitivities (A/A-GI_{50} > G/G-GI_{50}). For example, one of the topoisomerase II inhibitors for which the different genotypes of the YWHAQ SNP significantly differed was doxorubicin. Cells wild-type for p53 and A/A in genotype required 7-fold more doxorubicin to arrest growth compared with p53 wild-type cells G/G in genotype. Specifically, as seen in Figure 3.2.2C, the 17 cell lines wild-type for p53 required, on average, 0.13μmol/L of doxorubicin to arrest by 50%, and cells ranged from 0.01 to 0.64 μmol/L in their respective GI_{50} values. Those five cell lines that required the highest concentration (average 0.33 μmol/L) were significantly enriched in cells containing the A-allele compared with those five cell lines needing the least amount of agent (average 0.03 μmol/L, p=0.0002). By contrast, no significant differences in the A-allele frequencies were seen in the extreme groups of the p53 mutant cell lines (Figure 3.2.2C). Importantly, the allelic differences remain significant after performing a

multivariate analysis, including the tissue of origin and the mutational status of the other cancer-related genes for those cells with a wild-type p53 gene (p_{GG}=0.00004, p_{AA}=0.0013; Table 3.2.1[1]).

Table 3.2.1. Allelic differences for candidate SNP loci in growth responses to chemotherapeutics of NCI60 cells either wild type or mutant for p53.

			p53 mutant					p53 wild type				
			Univariate			Multivariate		Univariate			Multivariate	
Gene	SNP	GI50	p-value	No. of Significant Drugs	Fold ratio	Significant Genotypes	p-value	p-value	No. of Significant Drugs	Fold ratio	Significant Genotypes	p-value
YWHAQ	rs6734469	G<A	1.0E+00	3	1.8	AG, AA	1.3e-02, 2.4e-02	5.6E-47	108	5.4	GG, AA	4.0e-05, 1.3e-03
PPP2R2B	rs319227	A<C	8.9E-05	32	3.1	AC, CC	2.1e-02, 2.4e-02	1.3E-29	91	4.1	AA, CC	1.8e-02, 2.1e-02
PPP2R2B	rs319217	T<C	8.7E-04	33	3	CT, CC	2.1e-02, 2.4e-02	1.3E-29	91	4.1	TT, CC	1.8e-02, 2.1e-02
CCNG1	rs2069347	C<T	2.2E-03	28	2.5	CC, TT	1.0e-02, 1.1e-02	3.4E-26	87	5.7	TT, CT	2.0e-03, 2.4e-02
CD44	rs187115	T<C	3.9E-19	62	3.5	CC, TT	1.3e-03, 3.3e-03	2.1E-10	63	3.7	TT, CC	1.4e-03, 7.2e-03
PIAS1	rs1027154	G<C	1.0E+00	5	4.3	GG, CG	2.6e-03, 5.0e-03	1.7E-08	59	7.7	CG, GG	0.0e+00, 4.0e-05
KDR	rs2168945	G<T	4.8E-04	30	2.6	GG, TT	2.4e-03, 6.1e-03	1.2E-07	57	3.4	GG, GT	7.0e-04, 1.0e-02
CSE1L	rs2426127	T<C	1.0E+00	6	2	-	-	1.0E+00	6	1.9	-	-

[1] Alexei Vazquez

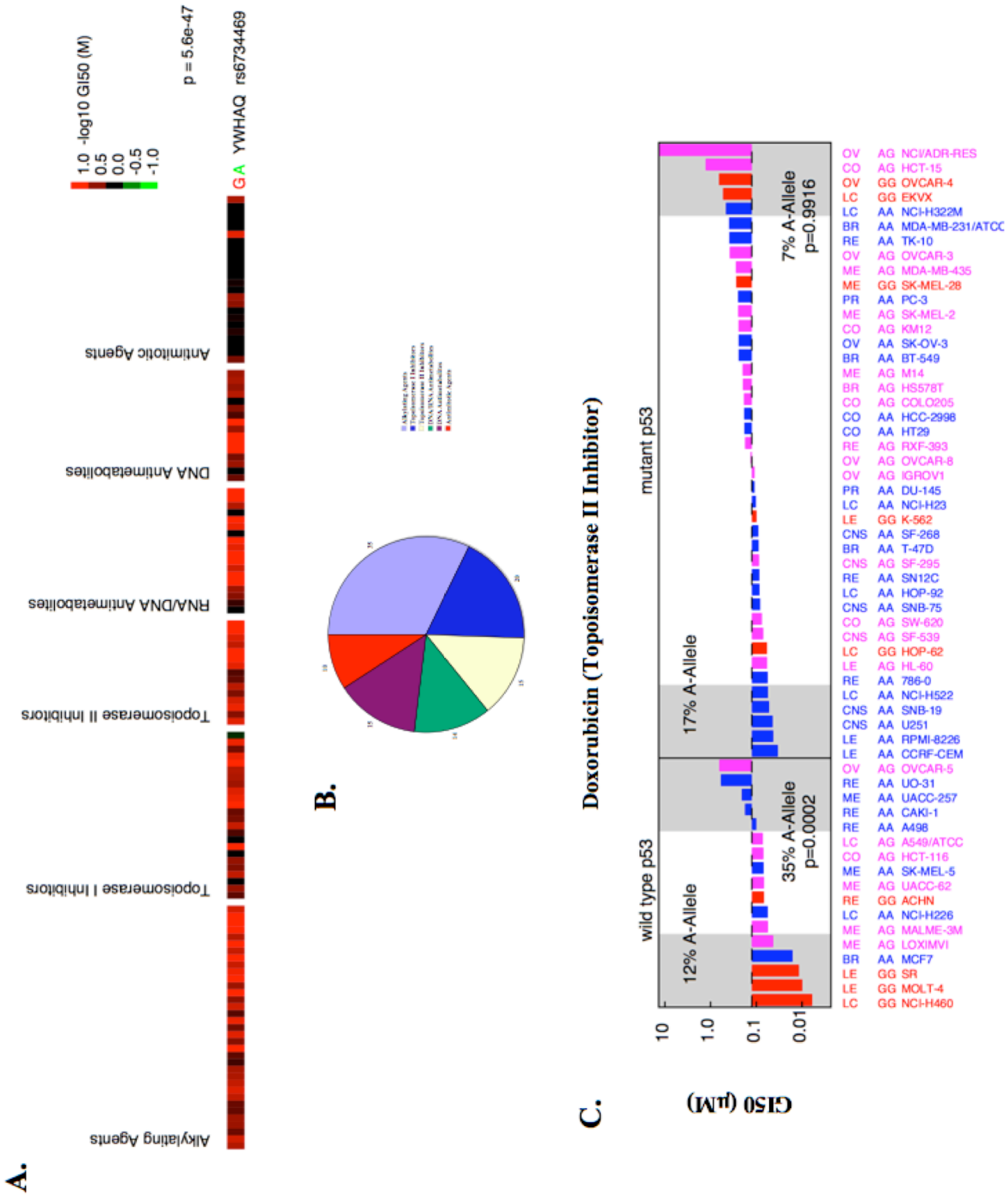


Figure 3.2.2. (legend on next page)

Figure 3.2.2 (previous page). The YWHAQ SNP (rs6734469) associates with significant allelic differences in cellular growth responses to chemotherapeutic agents in p53 wild-type cells. (A) A heat map summarizing the genotype-drug response associations for a panel of 132 chemotherapeutics in the NCI60 cell lines with wild-type p53. The color intensity is proportional to the difference between the mean $\log_{10}$ GI_{50} over cells with G/G and A/A homozygous genotypes, with red corresponding to lower GI_{50} (higher $-\log_{10}$ GI_{50}) in G/G than A/A containing cells and otherwise green. The genotypes of the YWHAQ SNP significantly differ in their growth response to 108 of the 132 agents tested. (B) A pie chart depicting the different classes of agents that the 108 drugs belong to based on their known mechanisms of action. (C) A graphical representation of the GI_{50} distributions of doxorubicin (a topoisomerase II inhibitor) across the NCI60 cell lines with those wild-type p53 (left) and those with mutant p53 (right). The five cell lines that required the highest concentration (average 0.33 μmol/L) were significantly enriched in cells containing the A allele compared with the five cell lines needing the least amount of agent (average 0.03 μmol/L, p=0.0002). In contrast, no significant differences in the A-allele frequencies were seen in the extreme groups of the p53 mutant cell lines[1].

Interestingly, not all of the re-genotyped SNPs retained the strong dependence of the wild-type p53 gene for the significant allelic differences in cellular growth responses to the standard chemotherapeutic agents (Table 3.2.1). For example, the allelic differences for the SNP in CD44, and to a lesser extent for the SNP in KDR, were seen with a similar number of agents and with similar strengths in both p53 wild-type and p53 mutant cells (Table 3.2.1). These data suggest that these allelic differences are not dependent on the mutational status of the p53 gene and that these two populations of cells can be grouped together in further analyses of these SNPs.

The allelic differences in cellular growth responses to the standard chemotherapeutic agents according to the SA univariate analysis, the associated multiple hypothesis test, and the multivariate test in all 59 cells regardless of the mutational status of the p53 gene are depicted in Table 3.2.2. Indeed, the strongest effects were seen for the SNP in the CD44 gene (rs187115). Specifically, the results of the univariate analysis suggest that cells with the different genotypes of the CD44 SNP significantly differ in

[1] Alexei Vazquez

their growth response to 96 of the 132 agents tested ($p=8.1\times10^{-24}$; Table 3.2.2 and Figure 3.2.3A). Significant allelic differences were observed with 21 alkylating agents, 24 and 14 topoisomerase I and II inhibitors, 17 and 8 RNA/DNA and DNA antimetabolites, and with 20 antimitotic agents (Figure 3.2.3A and B). Cells with the homozygote genotypes for CD44 SNP differed on average 2.8-fold in their respective drug sensitivities (C/C-GI_{50} > T/T-GI_{50}). For example, one of the topoisomerase II inhibitors for which the different genotypes of the CD44 SNP significantly differed in their growth response was also doxorubicin. Cells C/C in genotype required 10-fold more of the doxorubicin to arrest growth compared with cells T/T in genotype. Specifically, as seen in Figure 3.2.3C, those 10 cell lines that required the highest concentration (average 1.9μmol/L) were significantly enriched in cells containing the C allele compared with those 10 cell lines needing the least amount of agent (average 0.3μmol/L, p=0.0134). Importantly, for all 59 cell lines, the allelic differences remain significant after performing a multivariate analysis, including the tissue of origin and the mutational status of the other cancer-related genes, and the 264 p53 stress response SNP genotypes (p_{CC}=0.0000 and p_{TT}=0.0000; Table 3.2.2[1]).

Table 3.2.2. Allelic differences for candidate SNP loci in growth responses to chemotheraputics of NCI60 cell panel.

			Univariate			Multivariate	
Gene	**SNP**	**GI50**	**p-value**	**No. of Significant Drugs**	**Fold ratio**	**Significant Genotypes**	**p-value**
CD44	rs187115	T<C	8.1E-24	96	2.8	CC, TT	0.0e+00, 0.0e+00
PPP2R2B	rs319217	T<C	1.6E-05	63	2.4	CC, CT	5.4e-03, 2.2e-02
PPP2R2B	rs319227	A<C	1.6E-05	63	2.4	CC, AC	5.4e-03, 2.2e-02
YWHAQ	rs6734469_	G<A	1.4E-04	60	2.1	AA, AG	3.4e-04, 2.2e-02
KDR	rs2168945	G<T	2.8E-04	59	2.2	GG, TT, GT	0.0e+00, 2.1e-03, 7.6e-03
CCNG1	rs2069347	C<T	1.4E-02	52	2.7	TT, CC	2.1e-03, 1.5e-02
PIAS1	rs1027154	C<G	9.9E-01	29	3	GG, CG	4.5e-04, 2.2e-02
CSE1L	rs2426127	C<T	1.0E+00	27	2.3	CC	1.50E-002

[1] Alexei Vazquez

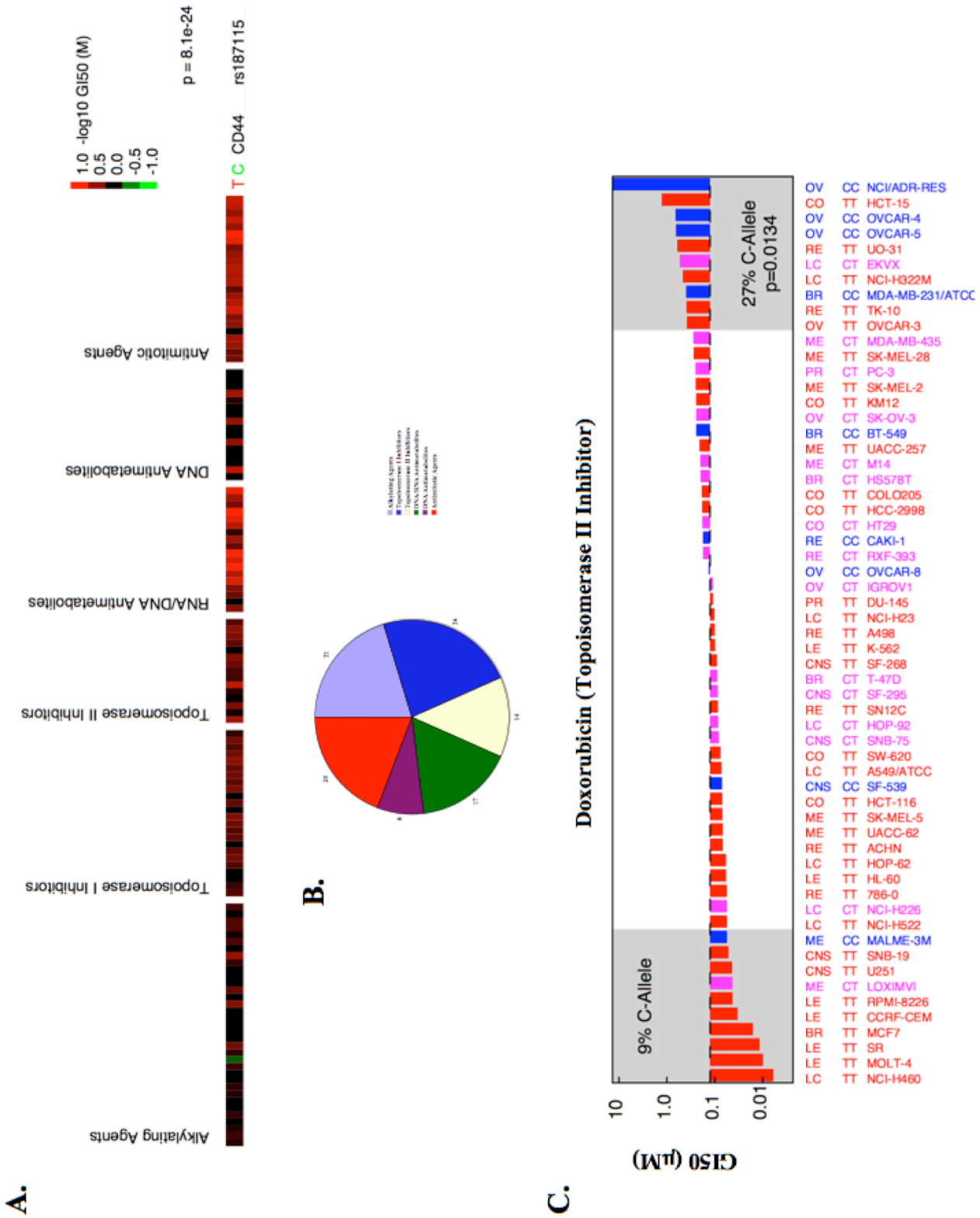


Figure 3.2.3. (legend on next page)

Figure 3.2.3. (previous page) The CD44 SNP (rs187115) associates with significant allelic differences in cellular growth responses to chemotherapeutic agents in cells. (A) A heat map summarizing the genotype-drug response associations for a panel of 132 chemotherapeutics with known mechanism of action in the NCI60 cell lines. The color intensity is proportional to the difference between the mean $\log_{10}$ GI_{50} over cells with T/T and C/C homozygous genotypes, with red corresponding to lower GI_{50} (higher $-\log_{10}$ GI_{50}) in T/T than C/C containing cells and otherwise green. The genotypes of the CD44 SNP significantly differ in their growth response to 96 of the 132 agents tested. (B) A pie chart depicting the different classes of agents that the nine drugs belong to based on their known mechanisms of action. (C) A graphical representation of the GI_{50} distributions of doxorubicin across the NCI60 cell lines. The 10 cell lines that required the highest concentration (average 1.9 μmol/L) were significantly enriched in cells containing the C-allele compared with the 10 cell lines needing the least amount of agent (average 0.3 μmol/L, $p=0.0134$)[1].

3.2.2.2. SNP-alleles with reduced chemosensitivities associate with worse sarcoma survival and earlier sarcoma incidence

The above-described allelic differences in cellular growth responses to chemotherapeutic agents for these SNPs suggest that individuals carrying the alleles that were associated with weaker growth responses to chemotherapeutics in cell culture could associate with poorer outcomes upon cancer onset due to a poorer response to therapies. To begin to test this, we analyzed these loci in 130 patients who were diagnosed with and treated for STS, a well-described, p53-surveilled tumor (Li, Strong et al. 1990; Donehower, Harvey et al. 1992; Momand, Jung et al. 1998). The YWHAQ and CD44 SNPs were successfully genotyped in 129 and 128 of the 130 patients, respectively. The genotype frequencies in the 129 STS patients of the CD44 SNP were 11.7% for the C/C genotype, 45.3% for the T/C genotype, and 43.0% for patients T/T in genotype (Table 3.2.3). The genotype distributions of the YWHAQ SNP were 27.1%, 47.3%, and 25.6%

[1] Alexei Vazquez

for the A/A, A/G, and G/G genotypes, respectively (Table 3.2.3). The allelic frequencies of both SNPs did not differ significantly from the frequencies of the controls of the same ethnic origin (Table 3.2.3).

Table 3.2.3. Genotype frequencies of CD44 and YWHAQ SNPs in cases and controls.

SNP	Genotype	Blood Donors n(%)	STS Patients n(%)
CD44 SNP	C/C	76 (15.3)	15 (11.7)
	C/T	216 (43.4)	58 (45.3)
	T/T	206 (41.4)	55 (43.0)
YWHAQ (14-3-3 tau) SNP	G/G	117 (23.5)	33 (25.6)
	A/G	242 (48.6)	61 (47.3)
	A/A	139 (27.9)	35 (27.1)

To assess the impact of both SNPs on patient outcomes, Cox multivariate regression analyses were performed, adjusting for known prognostic factors of STS: tumor stage, resection type (R-status), and gender (Molife, Lorigan et al. 2001; Lahat, Tuvin et al. 2008). Interestingly, the sarcoma patients harboring the alleles of the YWHAQ and CD44 SNPs, which in the NCI60 analysis were associated with weaker growth responses to chemotherapeutics, associated with poorer overall survival. Those patients homozygous for the C-allele of the CD44 SNP were associated with a 2.16-fold increased relative risk for tumor-related death compared with individuals C/T and T/T in genotype ($p=0.041$ Table 3.2.4). Furthermore, patients carrying either one or two copies of the A-allele of the YWHAQ SNP associated with a significantly worse prognosis compared with patients homozygous for the G allele ($p=0.043$; relative risk, 1.92; Table 3.2.4). As predicted by the results of the NCI60 analysis, the allelic differences in overall

survival became notably stronger when the survival analysis was restricted to those 83 patients who received radiotherapy/chemotherapy. Patients C/C for the CD44 SNP associated with a 2.89-fold increased risk for tumor-related death compared with patients C/T and T/T for CD44 who also received DNA-damaging therapies (p=0.011; Table 3.2.4, Figure 3.2.4A). Patients harboring either one or two copies of the A-allele of the YWHAQ SNP treated with DNA-damaging therapies associated with a 2.77-fold increased risk of tumor-related death compared with patients who are similarly treated but with a G/G genotype (p=0.011; Table 3.2.4, Figure 3.2.4B).

Table 3.2.4. Survival analysis in soft-tissue sarcoma patients

SNP	Patient Group	Genotype	n*	Relative risk**	p-value***
CD44 SNP	All patients	C/C	15	1	-
		C/T	58	0.44	0.044
		T/T	55	0.49	0.086
		C/T and T/T	113	0.46	0.041
	Radio-/chemotheraputically treated patients	C/C	13	1	-
		C/T	30	0.34	0.022
		T/T	39	0.35	0.022
		C/T and T/T	69	0.35	0.011
YWHAQ (14-3-3 tau) SNP	All patients	G/G	33	1	-
		A/G	61	2.2	0.021
		A/A	35	1.5	0.275
		A/G and A/A	96	1.9	0.043
	Radio-/chemotheraputically treated patients	G/G	25	1.00	-
		A/G	37	3.0	0.009
		A/A	21	2.6	0.040
		A/G and A/A	58	2.8	0.011

*Number of patients

** Relative risk of tumor-related death

***P-values were calculated using the Cox's multivariate regression analysis, adjusted to tumor stage, resection-type and gender

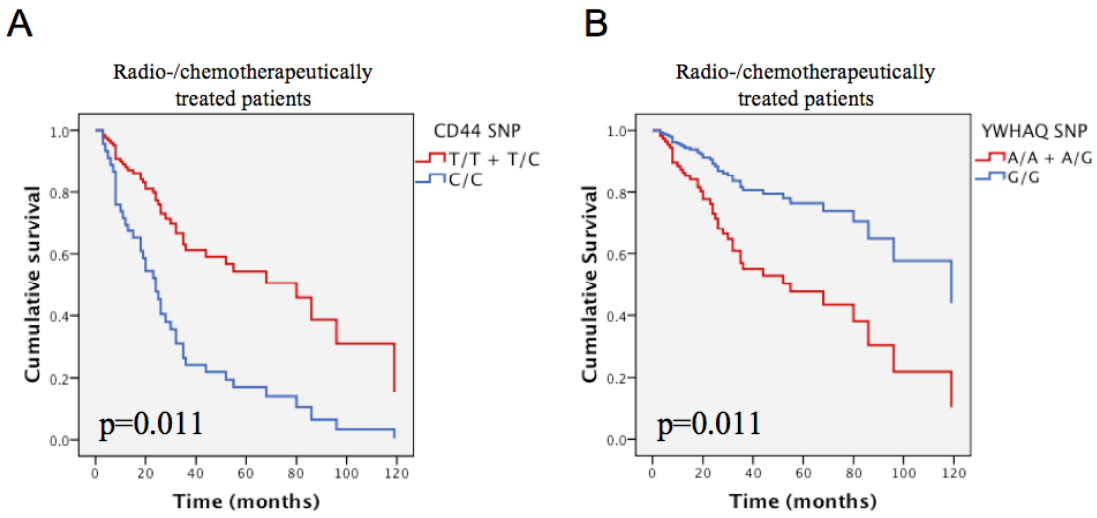


Figure 3.2.4. The alleles of the CD44 SNP and the YWHAQ SNP, which associate with a weaker apoptotic response to chemotherapeutic treatment in the NCI60 cell lines, also correlate significantly with poorer overall survival in STS patients. The graphs display the survival curves of radiotherapeutically and/or chemotherapeutically treated patients for the CD44 SNP genotypes (A) and for the YWHAQ SNP (B) plotted against the survival time in months. Chemotherapeutically treated patients C/C for the CD44 SNP associated with a 2.89-fold increased risk for tumor-related death compared with patients C/T and T/T (p=0.011) in genotype. Patients harboring either one or two copies of the A-allele of the YWHAQ SNP treated with DNA-damaging therapies associated with a 2.77-fold increased risk of tumor-related death compared with patients with a G/G genotype (p=0.011). All p-values were calculated with the Cox multivariate regression analysis adjusted to tumor stage, resection type (R-status), and gender.

The above-described results support a model that these SNPs affect cellular stress responses, thereby resulting in altered survival rates. As cellular stress responses like those mediated by the p53 pathway are also important in tumor suppression, we explored the effects of these loci on the age-dependent incidence of these tumors. Interestingly, the alleles of both SNPs, which in the NCI60 analysis associated with weaker growth responses to chemotherapeutics and, in the survival analysis, associated with decreased

survival rates, also associated with a significantly earlier age of tumor onset. Patients with a C/C genotype for the CD44 SNP were diagnosed on average at 49.1 years of age (range 16–83 years), patients C/T in genotype at 54.5 years (range 14–84 years), and patients T/T in genotype at 59.8 years of age (range 24–87 years; Table 3.2.5; p=0.002, Permutation test). Thus, the C/C homozygotes showed on average a 10.7-year earlier tumor onset than individuals homozygous for the T-allele (Table 3.2.5). Similar associations were observed for the YWHAQ SNP, whereby patients homozygous for the A-allele were diagnosed with a mean age of 52.8 years (range 17–85 years), patients A/G in genotype at 55.5 years (range 14–85 years), and patients G/G in genotype at 59.9 years (range 22–87 years; Table 3.2.5; p=0.006). Thus, patients with the A/A genotype showed on average a 7.1-year earlier age of onset than patients G/G in genotype (Table 3.2.5).

Table 3.2.5. Age of diagnosis analysis of CD44 and YWHAQ SNPs.

SNP	Genotype	n	Age of Diagnosis*	p-value**	Average Difference between Homozygouts
CD44 SNP	C/C	15	49.1 (53)	0.002	10.7 years
	C/T	58	54.5 (59)		
	T/T	55	59.8 (62)		
YWHAQ (14-3-3 tau) SNP	G/G	33	59.9 (59)	0.006	7.1 years
	A/G	61	55.5 (61)		
	A/A	35	52.8 (54)		

*Average (Median) in years

**Permutation test for the differences between the average age of diagnosis for each genotype

3.2.3. Discussion

In this report, we build upon, expand, and use a methodology to uncover functional variants in p53 stress response genes with data from the NCI60 human tumor cell line anticancer drug screen. We scanned 142 genes known to affect the p53 stress

response and identified seven SNPs in six genes that possess allelic differences in cellular growth responses to standard chemotherapeutic agents. The greatest differences in the p53 wild-type cells were observed for the YWHAQ SNP (rs6734469, $p=5.6\times10^{-47}$) and the strongest effects in all 59 cell lines, regardless of p53 mutational status, were seen for the CD44 SNP (rs187115, $p=8.1\times10^{-24}$).

The CD44 gene encodes a transmembrane glycoprotein involved in a vast range of cellular processes, such as regulation of growth and survival, differentiation, and motility (Ponta, Sherman et al. 2003; Naor, Wallach-Dayan et al. 2008). CD44 harbors tumor-promoting activities that include stimulating anchorage-independent cell growth and promote metastasis (Yu, Toole et al. 1997; Weber, Bronson et al. 2002; Barbour, Reeder et al. 2003). Furthermore, its expression is characteristic for breast and prostate cancer stem cells (Ponta, Sherman et al. 2003; Dou, Pan et al. 2007; Hurt, Kawasaki et al. 2008; Wright, Calcagno et al. 2008). Aberrated CD44 expression has also been reported to associate with tumor initiation and progression in various malignancies, such as colorectal, mammary, and prostate carcinomas, as well as neuroblastomas and sarcomas (Shtivelman and Bishop 1991; Wielenga, Heider et al. 1993; Yu, Toole et al. 1997; De Marzo, Bradshaw et al. 1998; Weber, Bronson et al. 2002). Certainly, the precise effects of altered CD44 expression on the initiation of human cancer and patients' prognosis need to be further elucidated, but the reported studies and our data suggest a functionally and clinically important role of CD44 and its polymorphic variants in human malignancies.

The YWHAQ gene encodes 14-3-3τ, a member of the 14-3-3-protein family. These proteins interact with many signaling pathways that are critical for apoptosis and cell proliferation, mainly through binding phosphoserine/threonine motifs (Morrison

2009). 14-3-3 binding induces conformational changes in the target protein that can alter its stability, catalytic activity, cellular localization, or susceptibility to intracellular proteases, kinases, and phosphatases (Tzivion, Gupta et al. 2006; Morrison 2009). The 14-3-3 proteins are highly conserved and consist of seven family members in mammals: the β, γ, ε, σ, ζ, τ, and η isoforms (Tzivion, Gupta et al. 2006). Interestingly, only YWHAQ (14-3-3τ) has been shown to promote apoptosis directly upon genotoxic stress (Waterman, Stavridi et al. 1998; Stavridi, Chehab et al. 2001; Wang, Liu et al. 2004). Furthermore, 14-3-3τ has also been shown to bind to wild-type p53 after ionizing radiation of cells and to stimulate transcriptional activity of p53 (Waterman, Stavridi et al. 1998; Stavridi, Chehab et al. 2001). This observation is consistent with the observed dependence of wild-type p53 for the allelic differences in the cellular growth responses to chemotherapeutics for the YWHAQ SNP in our study. Importantly, the A-allele, which associates with a significantly weaker drug response in the NCI60 panel, also associates with an early onset of disease and an increased risk of tumor-related death in STS patients, specifically in patients who received radiochemotherapeutic treatment (Figure 3.2.4).

The observations reported here remain to be validated in other patient cohorts and, importantly, the regulatory changes associated with YWHAQ and CD44 SNPs remain to be determined. As we will describe in more detail in Chapter 3.4, the identified YWHAQ SNP (rs6734469) is located in the second intron of the YWHAQ gene and the CD44 SNP (rs187115) in the first intron of CD44. Other SNPs are closely linked (r^2>0.8) to these two SNPs in the CEU population. Specifically, for the YWHAQ SNP, 24 SNPs in introns 2 to 5 and 3' of the gene are closely linked and, for the CD44 SNP, 7 SNPs in

intron 1 are linked (Chapter 3.4.2.1.1). To our knowledge, there have been no published attempts to characterize the regulatory regions of YWHAQ and no regulatory role of intron 1 of CD44 has been proposed. However, we will aim to begin to further elucidate the potential cellular mechanisms behind the significant allelic differences in cellular chemosensitivities, sarcoma incidence and survival in Chapter 3.4. Therein, we will present some evidence for CD44 SNP that those differences might be due to allele-specific regulation of CD44 transcript expression in cells. Taken together, the data presented in this Chapter support the model that both CD44 and 14-3-3τ play a significant regulatory role in cellular stress responses, thereby affecting sarcoma incidence and survival.

3.3. Gender- and age-dependent incidence of soft-tissue sarcoma screen identifies SNPs in the estrogen receptor alpha gene that interact with MDM2 SNP309 to affect human cancer

3.3.1. Introduction

3.3.2. Results

3.3.2.1. Gender- and age-dependent incidence of soft-tissue sarcoma screen identifies potential functional SNPs in the estrogen receptor alpha (ESR1) gene

3.3.2.2. The identified SNPs in the ESR1 gene interact with MDM2 SNP309 to affect cancer formation in females

3.3.3. Discussion

3.3.1. Introduction

The functional p53 pathway polymorphism MDM2 SNP309 in the enhancer region of the key negative regulator of p53 has been repeatedly shown to associate with allele- and gender-specific differences in tumor diagnosis in various malignancies (Bond, Hirshfield et al. 2006; Bond, Menin et al. 2006; Bartel, Jung et al. 2008). Interestingly, predominantly female carriers of the G-allele of MDM2 SNP309 have been shown to be diagnosed with cancers, such as soft-tissue sarcomas, lymphoma and colorectal carcinoma, significantly earlier than T-allele carriers (Bond, Hirshfield et al. 2006; Bond, Menin et al. 2006). For example, women homozygous for the G-allele of SNP309 showed a 13-year earlier lymphoma diagnosis than T/T women, while no significant differences in the age of tumor diagnosis were observed in men (Bond, Hirshfield et al. 2006). Interestingly, the greatest differences in the age-dependent incidence between the two genotypes were seen in women below 51 years of age, the average age of menopause, when female specific hormones like estrogen are at their highest levels (Bond, Hirshfield et al. 2006). Similar observations were made in sporadic STS, where a significantly earlier age of STS onset was only associated with the G/G women, and not men (Arva, Gopen et al. 2005). Specifically, women homozygous for the G-allele of SNP309 were diagnosed on average 14 years earlier than T/T women, further supporting the hypothesis that gender-specific hormones, like estrogen, could play a role in the ability of the SNP309 G-allele to accelerate tumor formation (Bond, Hirshfield et al. 2006). Interestingly, a similar interaction between MDM2 SNP309 and gender has been

recently reported for cancers associated with the Li-Fraumeni Syndrome, which is also characterized by a high incidence of soft-tissue sarcoma (Atwal, Rabadan et al. 2008; Wu, Krahe et al. 2011). Further support for this hypothesis came from the observation that an earlier age of tumor onset for G/G versus T/T women was greatest in the formation of breast cancers that expressed high levels of ER compared to tumors that expressed low levels of ER, suggesting that the G-allele of SNP309 can accelerate tumorigenesis only in the presence of an active estrogen-signaling pathway (Bond, Hirshfield et al. 2006).

Molecular and functional studies suggest that these gender-specific and estrogen receptor-dependent, clinical associations are the result of the transcriptional regulation of MDM2 gene expression by SP1 and its co-transcriptional factor estrogen receptor alpha (ERα) (Bond, Hu et al. 2004; Hu, Feng et al. 2007). ERα is activated by the primarily female-specific hormone estrogen, a critical component of the estrogen signaling pathway (Metivier, Penot et al. 2003). Previous studies had shown that the expression of ERα induces transcription of MDM2 (Kinyamu and Archer 2003; Hu, Feng et al. 2007) and many studies have reported that MDM2 mRNA and protein levels are higher in breast tumors, which express ERα (Sheikh, Shao et al. 1993; Gudas, Nguyen et al. 1995; Marchetti, Buttitta et al. 1995; Okumura, Saji et al. 2002). The regulation of MDM2 expression by estrogen is mediated, at least in part, through a well-characterized region of the MDM2-P2 promoter, in which SNP309 resides (Okumura, Saji et al. 2002; Kinyamu and Archer 2003; Phelps, Darley et al. 2003; Bond, Hu et al. 2004). Indeed, ERα has been shown to bind to this promoter region and activate transcription of the MDM2 gene (Metivier, Penot et al. 2003; Welboren, van Driel et al. 2009). A model was proposed that

enhancers with the G-allele have a higher affinity towards ERα-SP1-dependent activation of MDM2 transcription (Bond, Hu et al. 2004; Hu, Feng et al. 2007). Indeed, evidence to support this model was provided by the observations that estrogen activation leads to a preferential increase of MDM2 levels from the MDM2 SNP309 G-allele. This was shown by both allele-specific quantification of RNA polymerase loading on the MDM2 promoter region in a heterozygous cell line, MCF7, and in experiments carried out in a panel of ER-positive breast cancer cells (Hu, Feng et al. 2007).

Taken together, these observations have shown that MDM2 expression is regulated by ERα in an allele-specific manner, leading to heightened MDM2 levels specifically in cells carrying the SNP309 G-allele. These data suggest the possibility that the MDM2 SNP309 locus could alter the effects of estrogen on tumorigenesis, which results in the observed gender and age-dependent allelic differences in cancer incidence. Estrogen signaling manipulation has been shown to alter both the incidence and progression of cancer (Beckmann, Niederacher et al. 1997; Dietel, Lewis et al. 2005), and therefore women with a SNP309 G-allele could be affected differently by estrogen signaling manipulation than women with a T/T genotype. For example, it has been suggested that women with a G/G genotype and cancers which associate with gender-specific allelic differences in tumor risk, such as lymphoma, STS, and high ER-positive breast cancer would benefit from decreasing estrogen levels and this could significantly retard the progression of their disease (Bond, Hirshfield et al. 2006).

In this Chapter, we are interested in identifying other SNPs in the p53 network that could potentially identify those women that are more susceptible for the tumor promoting effects of estrogen signaling. To do this, we have designed a screening

approach to search for p53 network SNPs that, like MDM2 SNP309, associate with allelic differences in the age of tumor onset in a female-specific manner. The tumor type we have chosen is soft-tissue sarcoma. STS is clearly a p53-survailled tumor type (Li, Fraumeni et al. 1988; Malkin, Li et al. 1990; Toledo and Wahl 2006). Evidence to support this is the fact that STS is one of the most frequent cancers in humans and mice with a germline mutation in p53 (Li, Fraumeni et al. 1988; Malkin, Li et al. 1990; Donehower and Lozano 2009; Lozano 2010). Furthermore, the somatic amplification of MDM2 is found in approximately 30% of sporadic STS with wild-type p53 (Toledo and Wahl 2006). In addition, the female-specific allelic differences in the age of diagnosis for the MDM2 SNP309 locus have been noted for both hereditary and sporadic soft-tissue sarcomas (Bond, Hu et al. 2004; Bond, Hirshfield et al. 2006; Wu, Krahe et al. 2011).

In our screen, we utilize the same well-described cohort of German sporadic STS patients, in which the allelic differences of the MDM2 SNP309 locus were first described (Bond, Hu et al. 2004; Bond, Hirshfield et al. 2006). Interestingly, we identify a group of linked SNPs in the estrogen receptor alpha gene (ESR1) that associates with altered age-dependent onset of STS solely in females. We go on to demonstrate that the identified ESR1 SNPs significantly interact with MDM2 SNP309 to affect cancer formation in two independent STS cohorts and an ovarian cancer cohort, adding support to the hypothesis that SNPs in the p53 network can be utilized to more precisely identify patients at higher risk for cancer.

3.3.2. Results

3.3.2.1. Gender- and age-dependent incidence of soft-tissue sarcoma screen identifies potential functional SNPs in the estrogen receptor alpha (ESR1) gene

In order to begin to identify SNPs in p53 network genes that could also be regulated by estrogen signaling, we search for SNPs that, like MDM2 SNP309, associate with primarily female-specific differences in age-dependent incidence of STS. We focus our efforts on the above-described cohort of German sporadic STS patients (n=130) (Bond, Hu et al. 2004). Using a tag-SNP approach and a medium through-put Golden Gate genotyping platform, we genotype 639 common SNPs (MAF≥20%) that tag 591 out of 766 (77%) known haplotypes of 113 p53 network genes (Figure 3.3.1; Table 3.3.1). We determine the genotypes in a representative subset of 72 STS patients (30 males and 42 females) from the cohort. The 72 patients have been carefully selected to represent the entire cohort (130 patients) with respect to gender and age. Using the Jonckheere-Terpstra and Wilcoxon Tests, we determine the statistical significance of the noted differences in the mean age of tumor diagnosis between differing groupings of genotypes and alleles for each SNP in males and females: (i) A/A<A/B<B/B (ii) A/A<B (iii) A<B/B and (iv) A/A<B/B). Interestingly, the most significant, primarily female-specific, allelic differences in the age of sarcoma diagnosis are found with two SNPs in the ESR gene (rs2234693, p=0.00099 and rs9479130, p=0.00264; Figure 3.3.1; Table 3.3.2 and 3.3.3).

Table 3.3.1. The percent coverage of known haplotype blocks for each gene included in the analytical pipeline of the gender- and age dependent incidence of soft-tissue sarcoma screen.

Gene	Haplotype Blocks* (n)	Genotyped in Analysis (n)	Coverage (%)
APC	4	4	100.0
APEX1	2	2	100.0
BAD	2	2	100.0
BAX	2	2	100.0
BCAR1	3	3	100.0
CBL	3	3	100.0
CCNG1	2	2	100.0
CDKN1A	1	1	100.0
CDKN2A	2	2	100.0
CHEK2	6	6	100.0
CSE1L	4	4	100.0
CSNK2A1	4	4	100.0
CTNNB1	2	2	100.0
DUSP16	4	4	100.0
DYRK2	3	3	100.0
E2F1	1	1	100.0
E4F1	1	1	100.0
EP300	4	4	100.0
ESR2	6	6	100.0
FRAP1	2	2	100.0
GADD45A	2	2	100.0
GTF2H1	4	4	100.0
GTF2H3	2	2	100.0
GTF2H4	1	1	100.0
HIPK2	1	1	100.0
HRAS	1	1	100.0
ILK	3	3	100.0
LRDD	2	2	100.0
MAP4	2	2	100.0
MAPK8	3	3	100.0
MDM2	2	2	100.0
MMP2	5	5	100.0
MTA2	1	1	100.0
PCNA	1	1	100.0
PDPK1	1	1	100.0
PIAS1	4	4	100.0
PIK3CD	6	6	100.0
PIK3R2	1	1	100.0
PIK3R3	3	3	100.0
PPP1R13B	3	3	100.0
PTEN	1	1	100.0
RBBP7	2	2	100.0
RELA	1	1	100.0

SHC1	[1]1	1	100.0
TDG	4	4	100.0
TP53BP1	2	2	100.0
TSC2	1	1	100.0
VEGFA	3	3	100.0
KDR	13	12	92.3
TP63	71	63	88.7
ERCC2	8	7	87.5
TCF7L1	29	25	86.2
FAS	7	6	85.7
NUMB	7	6	85.7
MAP3K5	12	10	83.3
TP73	18	15	83.3
MAPK10	21	17	81.0
ITGB1	5	4	80.0
MAP2K4	10	8	80.0
MAPK9	15	12	80.0
RAF1	5	4	80.0
SOS2	5	4	80.0
SRC	5	4	80.0
TNFRSF10B	10	8	80.0
FLT1	28	22	78.6
ESR1	23	18	78.3
FHL2	9	7	77.8
EGFR	31	24	77.4
TCF7L2	31	24	77.4
CASP3	4	3	75.0
JMY	4	3	75.0
MAPK1	4	3	75.0
PIK3CA	4	3	75.0
RCHY1	4	3	75.0
EGF	7	5	71.4
LEF1	7	5	71.4
IGF1R	71	50	70.4
AKT1	3	2	66.7
AKT2	3	2	66.7
ATM	3	2	66.7
ATR	3	2	66.7
ERCC3	3	2	66.7
hCG_1789827	3	2	66.7
HDAC2	6	4	66.7
MDM4	3	2	66.7
PIK3CB	3	2	66.7
RFWD2	3	2	66.7
RHEB	6	4	66.7

Table 3.3.1 (continued)

TP53BP2	3	2	66.7
XPC	6	4	66.7
AKT3	5	3	60.0
FOXO3	5	3	60.0
MAP2K1	5	3	60.0
MAP2K2	10	6	60.0
IGF1	7	4	57.1
NFKB1	7	4	57.1
USP7	11	6	54.5
CASP9	4	2	50.0
CREBBP	8	4	50.0
PMAIP1	2	1	50.0
PXN	4	2	50.0
RB1	2	1	50.0
SOS1	2	1	50.0
TSC1	6	3	50.0
PIK3R1	21	10	47.6
APAF1	5	2	40.0
GSK3B	9	3	33.3
PTK2	9	3	33.3
REL	3	1	33.3
RELB	3	1	33.3
P2RY5	1	0	0.0
Total	766	591	77.2

Table 3.3.1. (continued) *HapMap Phase II data; MAF $\geq$ 20%

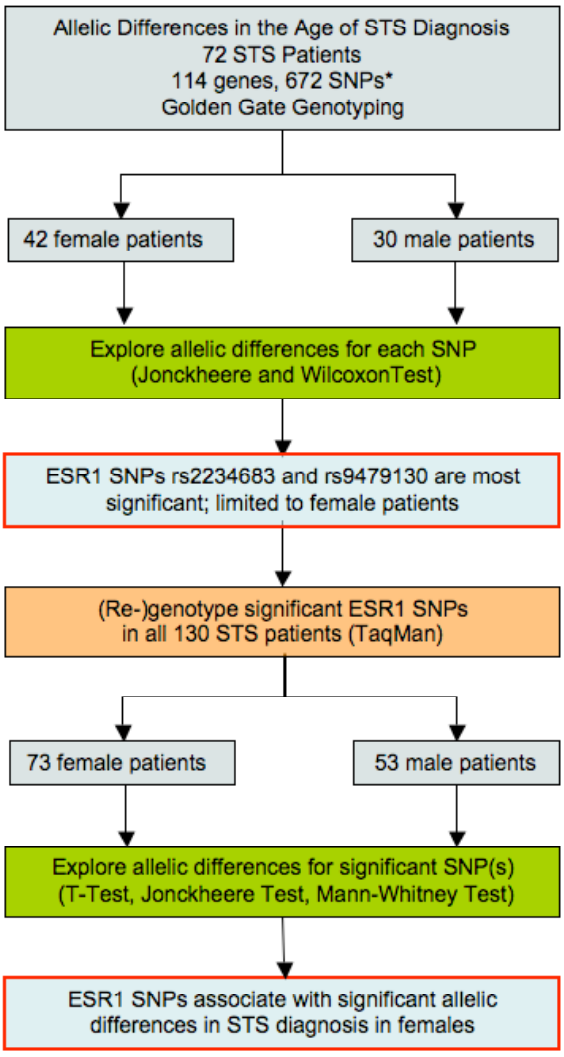


Figure 3.3.1. Flowchart depicting the design of the gender- and age-dependent incidence of soft-tissue sarcoma screen. Two closely linked SNPs in the ESR1 gene associated with the most significant differences in the age of STS diagnosis. * The SNPs (MAF ≥ 20%) were selected using a tag-SNP approach utilizing HapMap Phase II data, which resulted in 77.2% haplotype coverage in the studied genes.

Table 3.3.2. Top 15 SNP hits for females from the gender- and age dependent incidence of soft-tissue sarcoma screen. Abbreviations: Pvals_all, P-values all patients; Pvals_M, P-values males; Pvals_F, P-values females

SNP_no.	Gene symbol	Pvals_all	Pvals_M	Pvals_F
rs2234693	ESR1	0.251893034	0.073748221	0.000990995
rs9479130	ESR1	0.08693691	0.349586626	0.002638451
rs6424089	TP73	0.102357805	0.511889822	0.002939213
rs867431	IGF1R	0.028185802	0.332324764	0.004258443
rs1709183	ESR1	0.030851226	0.670505381	0.00441516
rs827423	ESR1	0.147587522	0.300464045	0.005025727
rs13112959	MAPK10	0.009191586	0.141288999	0.00509181
rs3789051	MDM4	0.005891465	0.568456614	0.005639935
rs7167580	IGF1R	0.001011361	0.038023299	0.006048884
rs3916874	ERCC2	0.001090044	0.103340444	0.006256587
rs7711939	JMY	0.07215003	0.375567557	0.00699532
rs4648546	TP73	0.055877346	0.876164629	0.008737007
rs2251687	IGF1R	0.001535988	0.164698123	0.010346977
rs2267922	PIK3R2	0.015623228	0.216849535	0.010979039
rs4259415	TNFRSF10B	0.429375765	0.028956449	0.011087895

Table 3.3.3. Top 14 SNP hits for males from the gender- and age dependent incidence of soft-tissue sarcoma screen. Abbreviations: Pvals_all, P-values all patients; Pvals_M, P-values males; Pvals_F, P-values females

SNP_no.	Gene_symbol	Pvals_all	Pvals_M	Pvals_F
rs8065164	MAP2K4	0.52565717	0.002797203	0.057998776
rs17418221	MAPK10	0.173474847	0.003216496	0.354927289
rs12507758	MAPK10	0.354309348	0.00441006	0.046215944
rs3025010	VEGFA	0.091007827	0.004649761	0.231892689
rs2715419	IGF1R	0.236062687	0.004749462	0.40798528
rs2684806	IGF1R	0.013597737	0.008393861	0.223401427
rs2684807	IGF1R	0.009525538	0.008422582	0.156897644
rs4273389	ATR	0.097684008	0.008587303	0.288437704
rs7656914	MAPK10	0.704269434	0.009291423	0.070066981
rs129968	CREBBP	0.042796063	0.009322428	0.613039118
rs2043228	TCF7L1	0.049458253	0.01053841	0.776502793
rs6823664	MAPK10	0.542483193	0.010822253	0.037435606
rs6709118	TCF7L1	0.270125962	0.011909938	0.474947665
rs1870583	MAP2K4	0.156258578	0.013469153	0.037001638

Next, we genotyped both ESR1 SNPs using very accurate TaqMan SNP genotyping assays (Applied Biosystems) in order to extend these findings to all 130 patients of the STS cohort, and to validate and complete the initial genotyping. Genotypes were obtained for almost all samples (9 missing genotypes for rs2234693 and 8 missing for rs9479130) and the genotypes of the SNPs were found to be 95% linked in the cohort, similar to what was found in the HapMap dataset for the Caucasian population (CEU, r^2=1). Of the 121 patients successfully genotyped for ESR1 SNP^{rs2234693}, 32 (T/T, 26.4%) were homozygous for the T-allele, 25 (C/C, 20.7%) were homozygous for the C-allele and 64 (T/C, 52.9%) were heterozygous (Table 3.3.4). Importantly and in line with the initial observations, females showed significant allelic differences in the age of STS diagnosis for both ESR1 SNPs, whereby women homozygous for the major alleles were diagnosed on average 12 years earlier than those homozygous for the minor alleles (p=0.0374; T-Test, Figure 3.3.2A, Table 3.3.5), and no significant differences between genotypes in the male patients were observed (Figure 3.3.2B, Table 3.3.5).

Table 3.3.4. ESR1 SNP rs2234693. Genotype frequencies in German STS cohort

	ESR1 SNP rs2234693			All genotypes
	C/C	T/C	T/T	
Total	25 (20.7%)	64 (52.9%)	32 (26.4%)	121 (100%)
Males	14 (26.4%)	30 (56.6%)	9 (17.0%)	53 (100%)
Females	11 (16.2%)	34 (50.0%)	23 (33.8%)	68 (100%)

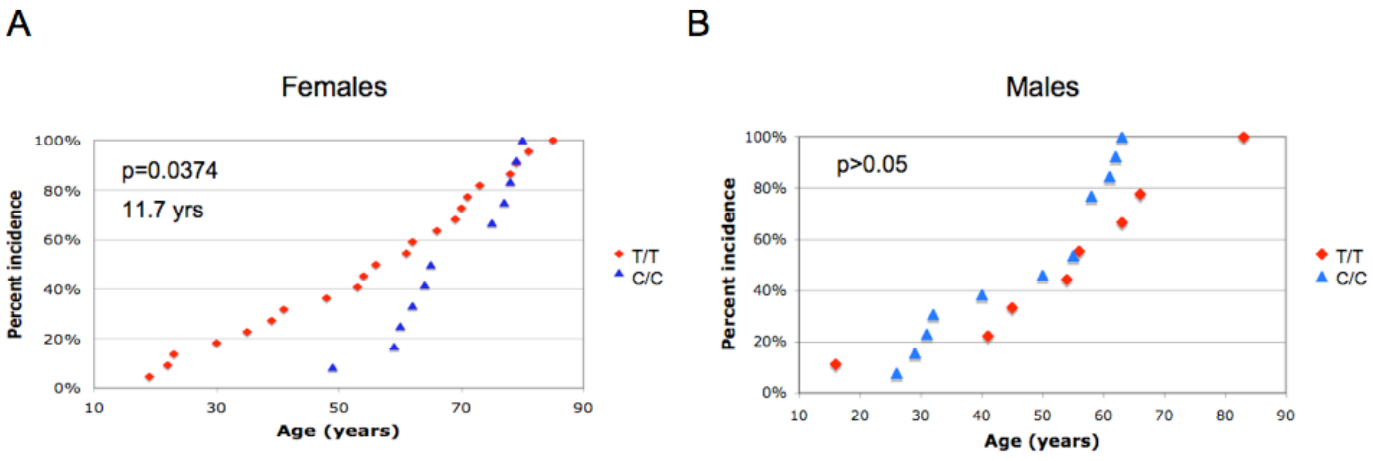


Figure 3.3.2. ESR1 SNP^{rs2234693} associates with allelic differences in the age of STS diagnosis in females, but not in males. (A) Female patients with a T/T genotype at the ESR1 SNP locus (red diamonds) were diagnosed on average 11.7 years earlier with STS than those homozygous for the C-allele (blue triangles), while females heterozygous for ESR1 SNP were intermediate (p=0.0374; T-Test). (B) No significant allelic differences in the age of STS diagnosis between individuals homozygous at the ESR1 SNP locus were observed in male patients

Table 3.3.5. ESR1 SNPrs2234693 Mean (median) age of STS diagnosis in years.

	Total	ESR1 SNPrs2234693			
	N	TT	CT	CC	P-value*
Total	121	54.9 (55.0)	57.4 (61.5)	55.7 (59.0)	>0.05
Males	53	51.6 (54.0)	57.7 (62.0)	46.0 (52.5)	>0.05
Females	68	56.3 (61.0)	57.3 (60.0)	68.0 (65.0)	0.0374

*T-test with Welch correction (TT : CC)

3.3.2.2. The identified SNPs in the ESR1 gene interact with MDM2 SNP309 to affect cancer formation in females.

As described above, estrogen has been shown to preferentially increase MDM2 levels from the SNP309 G-allele (Hu, Feng et al. 2007), and the regulation of MDM2 expression by estrogen has been shown to be mediated, at least in part, through a well-

characterized region of the MDM2-P2 promoter (Okumura, Saji et al. 2002; Kinyamu and Archer 2003; Phelps, Darley et al. 2003; Hu, Feng et al. 2007; Welboren, van Driel et al. 2009). We, therefore, reasoned that a polymorphic variant in the ERα gene that regulates estrogen signaling could interact with MDM2 SNP309 to affect cancer and therefore be utilized to more accurately predict an individual's cancer risk (Figure 3.3.3). In order to begin to explore potential functional interactions with MDM2 SNP309, we first explored the possibility that the above-noted allelic differences in the age of STS onset for the ESR1 SNPs could be restricted to female individuals of a specific genotype of MDM2 SNP309.

We successfully determined the genotypes of MDM2 SNP309 locus in 120 out of 121 STS patients that were genotyped successfully for ESR1 $SNP^{rs2234693}$. Of the 120 sarcoma patients genotyped for MDM2 SNP309, 55 (T/T, 45.8%) were homozygous for the T-allele, 17 (G/G, 14.2%) were homozygous for the G-allele and 48 (T/G, 40%) were heterozygous. We stratified the female patients according to their MDM2 SNP309 genotypes into two groups, one comprising G-allele carriers and one including those T/T in genotype, and analyzed the allelic differences in the age of STS diagnosis for ESR1 $SNP^{rs2234693}$ in each group (Table 3.3.6). Interestingly, all significant allelic differences for the ESR1 SNP were found in females harboring the G-allele of MDM2 SNP309 (Figure 3.3.4A). Specifically, female G-allele carriers homozygous for the major allele (T/T) were diagnosed on average 22.5 years earlier (32.5 years median difference) than female G-allele carriers homozygous for the minor allele (C/C) (p=0.031, Jonckheere-Terpstra Test, Figure 3.3.4A, Table 3.3.7). Interestingly, no significant allelic differences were observed for ESR1 SNP in females T/T in genotype for MDM2 SNP309 (Figure

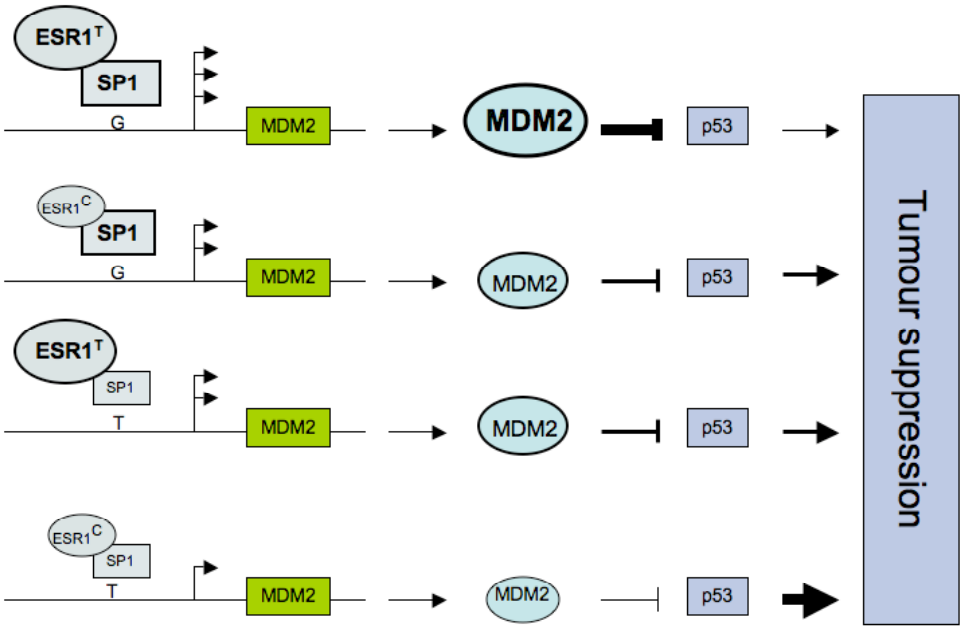


Figure 3.3.3. A model depicting the possible interaction between MDM2 SNP309 and ESR1 SNP^{rs2234693}, which might affect cancer risk through their effect on MDM2 transcription. The increased estrogen signaling associated with the T-allele of ESR1 SNP^{rs2234693} could lead to an increased transcription of MDM2 mRNA through interaction with its co-transcriptional activator SP1, specifically in cells harboring a MDM2 SNP309 G-allele. In contrast, in cells with an ESR1 SNP^{rs2234693} C/C genotype, which, upon speculation, have reduced estrogen signaling, or in cells homozygous for the T-allele of MDM2 SNP309, the activation of MDM2 expression through this mechanism would be less pronounced and result in higher p53 protein levels, more tumor suppression as well as decreased cancer risk.

3.3.4B, Table 3.3.7). No significant allelic differences were observed in the males upon similar patient stratification (data not shown). Together, these results suggest that the identified ESR1 SNPs interact with MDM2 SNP309 in a gender-specific manner to affect cancer risk, whereby the effects of the ESR1 SNPs are only found in female MDM2 SNP309 G-allele carriers.

Table 3.3.6. ESR1 SNP rs2234693. Genotype frequencies in German STS cohort

	ESR1 SNP rs2234693			All genotypes
	C/C	T/C	T/T	
MDM2 SNP309 G-allele				
Total	13 (20.0%)	35 (53.8%)	17 (26.2%)	65 (100%)
Males	6 (23.1%)	15 (57.7%)	5 (19.2%)	26 (100%)
Females	7 (17.9%)	20 (51.3%)	12 (30.8%)	39 (100%)
MDM2 SNP309 T/T				
Total	12 (21.8%)	29 (52.7%)	14 (25.5%)	55 (100%)
Males	8 (29.6%)	15 (55.6%)	4 (14.8%)	27 (100%)
Females	4 (14.3%)	14 (50.0%)	10 (35.7%)	28 (100%)

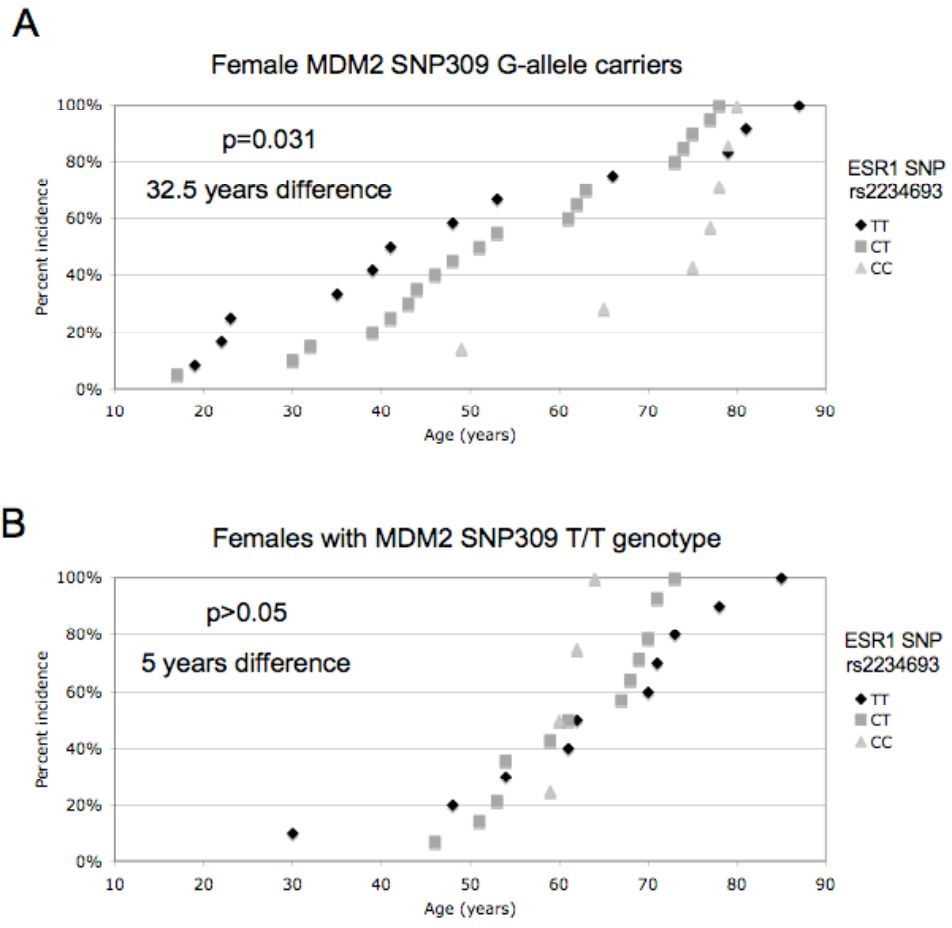


Fig 3.3.4. ESR1 SNP^{rs2234693} associates with allelic differences in the age of STS diagnosis in females harboring the G-allele of MDM2 SNP309. (A) Female patients with a T/T genotype at the ESR1 SNP locus (black diamonds) were diagnosed on average 22.5 years earlier (32.5-years median difference) with STS than those homozygous for the C-allele (light grey triangles), while females heterozygous for ESR1 SNP (grey squares) were intermediate (Jonckheere Test). (B) No significant allelic differences in the age of STS diagnosis between individuals homozygous at the ESR1 SNP locus were observed in female patients with a SNP309 T/T genotype.

Table 3.3.7. ESR1 SNP rs2234693 and the analysis of the age of STS diagnosis in female German patients, stratified according the MDM2 SNP309 genotype. Mean (median) age of diagnosis in years.

	Total	ESR1 SNP rs2234693			P-value*
	n	C/C	T/C	T/T	
MDM2 SNP309 G-allele	39	71.9 (77.0)	54.0 (52.0)	49.4 (44.5)	0.031
MDM2 SNP309 T/T	28	61.3 (61.0)	61.9 (64.0)	63.2 (66.0)	>0.05

*Junckheere-Terpstra Test T/T vs T/C vs C/C

To further validate the findings and exclude the possibility that the results are sample or population specific, we have explored these associations in a second and independent cohort of 167 Australian sarcoma patients (95 males, 72 females, Chapter 2.2.2, Table 2.10). This cohort comprised 95 soft-tissue sarcomas, 48 primary bone sarcomas and 24 sarcomas with undefined tissue origin (Chapter 2.2.2, Table 2.10). The Australian STS spectrum closely resembled the one in the German STS cohort (Chapter 2.2.1, Table 2.9). For example, among the most frequent tumor types were Liposarcomas (21.1% of the STS in the Australian and 23.1% in the German cohort) and Leiomyosarcoma (21.1% of the STS in the Australian and 17.7% in the German cohort). Most prominently, differences in the tumor spectrum were observed in the diagnosis malignant fibrous histiocytoma (MFH; 25.4% in German cohort, only 1 patient in the Australian cohort). However, these differences are most likely to be caused by recent changes in sarcoma classification. The German cohort was recruited in the years 1991-2001, when the classification MFH was commonly used for otherwise un-specified STS, a classification later replaced by the term non-otherwise specified (NOS) sarcoma (Matushansky, Charytonowicz et al. 2009). In the Australian cohort, NOS sarcoma make up 10.5% of the tumors classified as STS (none have such classification in the German cohort). In addition, it is likely that many of the 24 tumors, which are neither classified as STS nor as primary bone tumors in the Australian database, are in fact NOS soft-tissue

sarcoma. Therefore, the German and Australian databases are likely to also have a similar percentage of NOS/MFH tumors.

The ESR1 SNP^{rs2234693} and MDM2 SNP309 genotypes were determined successfully in 165 out of 167 Australian sarcoma patients. 57 (T/T, 34.5%) were homozygous for the T-allele of ESR1 SNP^{rs2234693}, 40 (C/C, 24.2%) were homozygous for the C-allele and 68 (T/C, 41.2%) were heterozygous (Table 3.3.8). 62 individuals (37.6%) were homozygous for the T-allele of MDM2 SNP309, 28 (17.0%) were homozygous for the G-allele and 75 patients (45.5%) were heterozygous. Female patients carrying the T-allele of ESR1 SNP^{rs2234693} associated with a non-significant, on average 5.3-year earlier age of sarcoma diagnosis than those C/C in genotype (p=0.4016, Mann-Whitney Test, Table 3.3.9), while a 1.9-year difference between the genotypes was observed in the males (p=0.3169 Mann-Whitney Test, Table 3.3.9). Interestingly, when patients were stratified according to the genotype at the MDM2 SNP309 locus (Table 3.3.8), significant gender-dependent allelic differences in the age of sarcoma diagnosis were observed for the ESR1 SNP^{rs2234693} locus. Similar to the observations made in the German STS cohort, female G-allele carriers for MDM2 SNP309, who harbor the T-allele of ESR1 SNP^{rs2234693}, were diagnosed on average 16.1-years earlier in life than G-allele carriers, C/C in genotype for the ESR1 SNP^{rs2234693} (p=0.0207, Mann-Whitney Test, two-sided; Table 3.3.9). In female T/T carriers for MDM2 SNP309, an opposite non-significant trend was noted, whereby female sarcoma patients carrying the T-allele of ESR1 SNP^{rs2234693} were diagnosed on average 14.6-years later than those C/C in genotype (p=0.1410, Mann-Whitney Test, two-sided; Table 3.3.9). No significant associations were noted in males of any genotype in any combination (Table 3.3.9). Similar results

were obtained in all patient groupings when the analysis was restricted to those 95 Australian patients who were diagnosed with STS. Hereby, female G-allele carriers who harbored the T-allele of ESR SNP^{rs2234693} were diagnosed on average 17.5 years earlier than female G-allele carriers homozygous for the minor allele (C/C) (p=0.0485, Mann-Whitney Test, one-sided). Together, these results further support the hypothesis that the identified ESR1 SNPs interact with MDM2 SNP309 in a gender-specific manner to affect sarcoma risk, whereby the effects of the ESR1 SNPs are only found in female G-allele carriers.

Table 3.3.8. ESR1 SNP rs2234693. Genotype frequencies in Australian sarcoma cohort

	ESR1 SNP rs2234693			All genotypes
	C/C	T/C	T/T	
MDM2 SNP309 G-allele and T/T				
Total	40 (24.2%)	68 (41.2%)	57 (34.5%)	165 (100%)
Males	25 (26.9%)	37 (39.8%)	31 (33.3%)	93 (100%)
Females	15 (20.8%)	31 (43.1%)	26 (36.1%)	72 (100%)
MDM2 SNP309 G-allele				
Total	25 (24.3%)	41 (39.8%)	37 (35.9%)	103 (100%)
Males	15 (25.0%)	23 (38.3%)	22 (36.7%)	60 (100%)
Females	10 (23.3%)	18 (41.9%)	15 (34.9%)	43 (100%)
MDM2 SNP309 T/T				
Total	15 (24.2%)	27 (43.5%)	20 (32.3%)	62 (100%)
Males	10 (30.3%)	14 (42.4%)	9 (27.3%)	33 (100%)
Females	5 (17.2%)	13 (44.8%)	11 (37.9%)	29 (100%)

Table 3.3.9. ESR1 SNP rs2234693 and the analysis of the age of STS diagnosis in Australian patients, stratified according the MDM2 SNP309 genotype. Mean (median) age of diagnosis in years.

	Total	ESR1 SNP rs2234693				P-value*
	n	C/C	T/C	T/T	T/T + T/C	
MDM2 SNP309 G-allele and T/T						
Total	165	48.7 (52.5)	43.9 (45.0)	45.4 (48.0)	44.5 (46.0)	>0.05
Males	93	47.6 (53.0)	42.2 (45.0)	45.7 (47.0)	43.8 (46.0)	>0.05
Females	72	50.6 (49.0)	45.9 (45.0)	44.9 (48.5)	45.4 (48.0)	>0.05
MDM2 SNP309 G-allele						
Total	103	52.2 (54.0)	40.0 (38.0)	44.4 (46.0)	42.1 (43.5)	0.0219
Males	60	48.1 (53.0)	40.8 (43.0)	43.2 (43.0)	42.0 (43.0)	>0.05
Females	43	58.3 (59.0)	38.9 (38.0)	46.2 (49.0)	42.2 (44.0)	0.0207
MDM2 SNP309 T/T						
Total	62	42.9 (37.0)	49.7 (54.0)	47.1 (50.0)	48.6 (52.0)	>0.05
Males	33	46.7 (53.0)	44.4 (47.5)	51.9 (58.0)	47.3 (48.0)	>0.05
Females	29	35.2 (27.0)	55.5 (57.0)	43.1 (48.0)	49.8 (54.0)	>0.05

* Mann-Whitney Test, two-sided T/T+T/C vs C/C

In order to further test this hypothesis, we were interested in exploring the interaction of these SNPs in a tumor type, wherein MDM2 SNP309 associated with estrogen receptor-dependent allelic differences in the age of diagnosis (Bartel, Jung et al. 2008). To do this, we genotyped ESR1 SNP^{rs2234693} in a cohort of 94 ovarian cancer patients from Germany (Chapter 2.2.3, Table 2.11), wherein patients with the G-allele of MDM2 SNP309 were shown to associate with a significantly 8.5-year earlier diagnosis (Bartel, Jung et al. 2008)) but only if their tumors expressed ER. 19 (T/T, 20.2%) were homozygous for the T-allele of ESR1 SNP^{rs2234693}, 28 (C/C, 29.8%) were homozygous for the C-allele and 47 (T/C, 50.0%) were heterozygous (Table 3.3.10). We utilized the previously determined MDM2 SNP309 genotypes, where 39 individuals (41.5%) were homozygous for the T-allele of MDM2 SNP309, 7 (7.4%) were homozygous for the G-allele and 48 patients (51.1%) were heterozygous. Of the 94 ovarian cancer patients, 52 had ER positive tumors and 42 had ER-negative tumors. In the patients with ER positive tumors, a non-significant, average 1.8-year earlier age of diagnosis was observed for patients carrying a T-allele at the ESR1 SNP^{rs2234693} locus compared to those C/C in genotype ($p>0.05$, Mann-Whitney Test, Table 3.3.11), while a 1.3-year difference between the genotypes was observed in the patients with ER-negative tumors ($p>0.05$; Mann-Whitney Test, Table 3.3.11).

However, when the 52 patients with ER-positive tumors were stratified on their MDM2 SNP309 genotypes and then further stratified on their ESR1 SNP^{rs2234693} genotypes, significant allelic differences were seen. Specifically, and similar to the observations made in both sarcoma cohorts, G-allele carriers for MDM2 SNP309 that harbor the T-allele for ESR1 SNP^{rs2234693} showed a 7.3-year average earlier age of onset

of ovarian carcinoma compared with G-allele carriers C/C in genotype (p=0.0477, Mann-Whitney-Test, one-sided; Table 3.3.11). No significant allelic differences were observed for ESR1 SNP^{rs2234693} in women with a SNP309 T/T genotype (Table 3.3.11). Furthermore, no significant differences for either SNP (or combination of SNPs), was noted in patients with-ER negative tumors (Table 3.3.11). Together, these results along with the results from the two sarcoma cohorts, support the hypothesis that the identified ESR1 SNPs interact with MDM2 SNP309 in a gender-specific and estrogen receptor-dependent manner to affect human cancer risk, whereby the effects of the ESR1 SNPs are only found in female G-allele carriers and/or tumors that express ER.

Table 3.3.10. ESR1 SNP rs2234693. Genotype frequencies in German ovarian cancer cohort

	ESR1 SNP rs2234693			All genotypes
	C/C	T/C	T/T	
All patients	28 (29.8%)	47 (50.0%)	19 (20.2%)	94 (100%)
ER positive tumours				
MDM2 SNP309 G-allele and T/T	15 (28.8%)	26 (50.0%)	11 (21.1%)	52 (100%)
MDM2 SNP309 G-allele	7 (23.3%)	18 (60.0%)	5 (16.7%)	30 (100%)
MDM2 SNP309 T/T	8 (36.4%)	8 (36.4%)	6 (27.3%)	22 (100%)
ER negative tumours				
MDM2 SNP309 G-allele and T/T	13 (31.0%)	21 (50.0%)	8 (19.0%)	42 (100%)
MDM2 SNP309 G-allele	6 (24.0%)	13 (52.0%)	6 (24.0%)	25 (100%)
MDM2 SNP309 T/T	7 (41.2%)	8 (47.1%)	2 (11.8%)	17 (100%)

Table 3.3.11. ESR1 SNP rs2234693 and the analysis of the age of ovarian cancer diagnosis in German patients, stratified according the MDM2 SNP309 genotype and estrogen receptor (ER) status. Mean (median) age of diagnosis in years.

	Total	ESR1 SNP rs2234693				P-value*
	n	C/C	T/C	T/T	T/T + T/C	
ER positive tumours						
MDM2 SNP309 G-allele and T/T	52	67.9 (67.0)	65.9 (66.5)	66.5 (62.0)	66.1 (66.0)	>0.05
MDM2 SNP309 G-allele	30	70.9 (71.0)	63.7 (65.0)	63.4 (60.0)	63.6 (65.0)	0.0477
MDM2 SNP309 T/T	22	65.3 (64.0)	70.9 (72.0)	69.2 (69.0)	70.1 (72.0)	>0.05
ER negative tumours						
MDM2 SNP309 G-allele and T/T	42	61.7 (61.0)	62.2 (64.0)	65.0 (62.5)	63.0 (63.0)	>0.05
MDM2 SNP309 G-allele	25	66.0 (63.0)	59.6 (57.0)	62.5 (64.0)	61.6 (62.0)	>0.05
MDM2 SNP309 T/T	19	61.0 (59.0)	66.4 (66.0)	62.0 (62.0)	65.5 (64.0)	>0.05

* Mann-Whitney-Test, one-sided T/T+T/C vs C/C

3.3.3. Discussion

In this Chapter, we have utilized a known clinical feature of MDM2 SNP309, specifically its association with female specific allelic differences in the age-dependent risk for various cancers, such as soft-tissue sarcoma, lymphoma and colorectal carcinoma (Bond, Hirshfield et al. 2006; Bond, Menin et al. 2006), to search for other polymorphisms in the p53 network that could also possibly affect cancer. We have genotyped 639 SNP in 113 p53 network genes and found that the closely linked SNPs rs2234693 and rs9479130 in the ESR1 gene associate with allelic differences in the age of STS diagnosis only in females. Hereby, we observed an 11.7-years earlier age of tumor diagnosis in females T/T in genotype for ESR1 $SNP^{rs2234693}$ compared to those homozygous for the C-allele ($p=0.0374$, Figure 3.3.2, Table 3.3.5). As we will describe in more detail in Chapter 3.4, these SNPs are closely linked to 14 other SNPs in this gene. Of the 16 SNPs, one is a synonymous SNP in exon 1 and 15 SNPs are found in introns 1 and 2 of the ESR1 gene (Chapter 3.4.2.1.1, Table 3.4.5; Figure 3.4.4). Interestingly, previous reports have correlated these SNPs to allelic differences in breast and endometrial cancer risk (Wedren, Lovmar et al. 2008; Anghel, Narita et al. 2010), whereby, similar to our findings, the major alleles associate with a higher risk for those malignancies. Moreover, allelic differences for this haplotype have also been seen for altered risk for other human conditions linked to estrogen signaling such as osteoarthritis (Riancho, Garcia-Ibarbia et al.), myocardial infarction (Schuit, Oei et al. 2004), schizophrenia (Weickert, Miranda-Angulo et al. 2008), differences in body height

(Schuit, van Meurs et al. 2004) and bone mass density (van Meurs, Schuit et al. 2003). One of these studies provided data that suggests that one or more of the SNPs could affect ESR1 transcript levels. Specifically, the major alleles of the SNPs were found to associate with higher ESR1 transcript levels in human brain tissues (Weickert, Miranda-Angulo et al. 2008).

ERα is a transcription factor and is known to positively regulate the MAPK (Migliaccio, Di Domenico et al. 1996), JNK (Hall, Couse et al. 2001) and PI3K/AKT pathways (Simoncini, Hafezi-Moghadam et al. 2000) and to regulate the signaling of transmembrane growth factor receptors such as EGFR, HER2, and insulin-like growth factor receptor I (IGFR1) through direct interaction with the receptors (Pietras and Marquez-Garban 2007). Importantly, ERα has also been shown to act as a modifier of the p53 network, through up-regulation of MDM2 expression (Okumura, Saji et al. 2002; Kinyamu and Archer 2003; Phelps, Darley et al. 2003; Hu, Feng et al. 2007; Welboren, van Driel et al. 2009), as well as through direct binding to p53 and repression of its transcriptional function, which results in inhibition of p53-mediated cell cycle arrest and apoptosis (Liu, Konduri et al. 2006; Sayeed, Konduri et al. 2007; Liu, Ip et al. 2009; Konduri, Medisetty et al. 2010). An up-regulation of ERα expression through the major alleles of the identified ESR1 haplotype could therefore lead to a heightened proliferation in cells and an increased cancer risk through one of the above mechanisms. We will aim to explore this model with functional and molecular assays and discuss it in detail in Chapters 3.4 and 3.5. Indeed, we will provide some evidence that this model might at least partly explain the observed clinical associations in support of the notion that the major allele of one or more SNP(s) from the identified ESR1 haplotype might indeed lead

to a relatively high ESR1 mRNA expression and protein levels, thus leading to a heightened proliferation in cells and affect cancer risk.

Based upon the clinical associations of MDM2 SNP309, whereby the G-allele associates with an increased risk for early tumor onset particularly in females, and molecular evidence that ERα up-regulates the expression of MDM2 preferentially through the G-allele (Kinyamu and Archer 2003; Bond, Hu et al. 2004; Hu, Feng et al. 2007; Welboren, van Driel et al. 2009), we were interested in exploring the possibility that the identified ESR1 haplotype can interact with MDM2 SNP309 to affect cancer susceptibility. Indeed, we observed allelic differences in the age of STS diagnosis at the ESR1 $SNP^{rs2234693}$ locus in female SNP309 G-allele carriers, while no effect was observed in females T/T for MDM2 SNP309. Specifically, women carrying the G-allele of MDM2 SNP309, who are homozygous for the T-allele of ESR1 $SNP^{rs2234693}$, were diagnosed with STS on average 22.5 years (or 19.6 years when carrying either one or two T-alleles) earlier than those carrying the G-allele of MDM2 SNP309 and C/C in genotype at the ESR1 $SNP^{rs2234693}$ locus (Figures 3.3.4 and 3.3.5, Table 3.3.7). Importantly, these findings were further supported and validated by similar observations in another sarcoma cohort from Australia, whereby in females who harbor the MDM2 SNP309 G-allele, patients carrying a T-allele of ESR1 $SNP^{rs2234693}$ associated with a 16.1-year earlier age of tumor diagnosis compared to those C/C in genotype ($p=0.0207$, Figure 3.3.5, Table 3.3.9). In addition, we observed the same significant trend in ovarian carcinoma patients who carry an MDM2 SNP309 G-allele and whose tumors express the estrogen receptor, with a 7.3-year allelic difference between females carrying the T-allele of ESR1 $SNP^{rs2234693}$ and C/C in genotype ($p=0.0477$, Figure 3.3.5, Table 3.3.11).

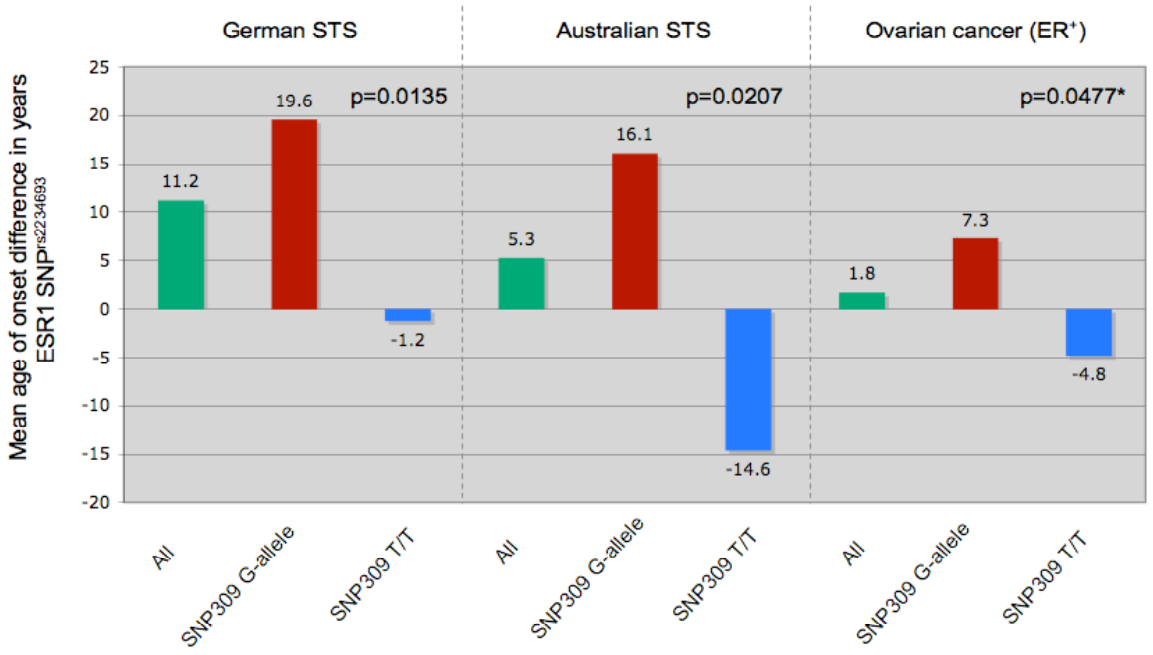


Figure 3.3.5. Mean age of cancer onset difference in years between female ESR1 SNP^{rs2234693} T-allele carriers and women with a C/C genotype, in all female patients and stratified according to the MDM2SNP309 genotype. In women carrying the G-allele of SNP309, a difference in the age of STS diagnosis is observed between females carrying the T-allele at the ESR1 SNP locus and women with a C/C genotype in a German (19.6 y difference, p=0.0135) and Australian (16.1 y difference, p=0.0207) STS cohort and in ovarian carcinoma patients (7.3 y difference in patients with ER-positive tumors, p=0.0477). In contrast, no significant differences at the ESR1 SNP locus were observed in the age of tumor diagnosis in females with a SNP309 T/T genotype. All p-values were calculated using a two-sided Mann-Whitney Test, except where indicated by an asterisk (one-sided Mann-Whitney test).

Taken together, the results suggest that polymorphic variants in the ERα gene could indeed interact with MDM2 SNP309 to affect cancer in females and therefore be utilized to more accurately predict an individual's cancer risk. Indeed, detailed molecular and functional analyses have provided compelling evidence that the co-transcriptional factor of ERα, SP1, binds with a higher affinity to the G-allele of MDM2 SNP309,

thereby up-regulating MDM2 expression in an allele-specific manner (Kinyamu and Archer 2003; Bond, Hu et al. 2004; Hu, Feng et al. 2007; Welboren, van Driel et al. 2009). It can be hypothesized that the proposed up-regulation of ESR1 mRNA and protein expression through a major allele of one or more SNPs from the identified ESR1 haplotype might result in the activation of the MDM2-P2 promoter specifically in cells with the G-allele of MDM2 SNP309 through the above-described molecular mechanism (Bond, Hu et al. 2004; Hu, Feng et al. 2007) (Figure 3.3.3). Subsequently, the heightened MDM2 expression could lead to an attenuation of p53 tumor suppression and result in increased cancer formation. In contrast, in cells with an ESR1 $SNP^{rs2234693}$ C/C genotype, which, upon speculation, have lower ERα levels, or in cells homozygous for the T-allele of MDM2 SNP309, the activation of MDM2 expression through this mechanism would be less pronounced and result in higher p53 protein levels, higher tumor suppressive activity and a decreased cancer risk (Figure 3.3.3). Indeed, in Chapters 3.4 and 3.5, we will provide some evidence that this model might at least partly explain the clinical associations observed in the cancer cohorts in support of the model that the identified ESR1 haplotype and MDM2 SNP309 might interact to affect cancer risk.

Estrogen signaling is commonly medicinally targeted in humans for a variety of reasons, such as contraception, infertility, relief of menopausal symptoms, as well as cancer prevention and treatment (Jordan 2006). Importantly, the manipulation of estrogen signaling has been shown to significantly affect cancer incidence as well as progression (Beckmann, Niederacher et al. 1997; Dietel, Lewis et al. 2005; Jordan 2006). This study suggests that women who carry the G-allele of MDM2 SNP309 and the T-allele of ESR1 $SNP^{rs2234693}$ could be affected differently by estrogen signaling manipulation than women

with other genotype combinations at these loci. Specifically, in both sarcoma cohorts 65 out of 139 female sarcoma patients carried this combination of SNP^{rs2234693} and SNP309 genotypes and were diagnosed at a median age of 44-49.5 years, which is below the average age of menopause, when estrogen levels are at their highest. These observations suggest that these females might benefit from decreasing estrogen levels, which could significantly retard the progression of their disease. Indeed, the selective estrogen receptor antagonist (SERM) Tamoxifen is widely used in the therapy of ER-positive invasive breast carcinoma, as it significantly reduces cancer progression (Jordan 2006). It will be interesting to study if reduction of estrogen signaling in cancer patients, such as soft-tissue sarcoma, might also be considered as a therapeutic option for women with a combination of the MDM2 SNP309 G-allele and ESR1 SNP^{rs2234693} T-allele, as the results presented here suggest.

3.4. The identification of candidate SNPs that could affect p53 network gene expression

3.4.1. Introduction

3.4.2. Results

3.4.2.1. The identification of potential causal SNPs that could underlie cellular and clinical phenotypes

3.4.2.1.1. Analysis of the linkage disequilibrium between polymorphisms in the MDM4, PPP2R5E, CD44, YWHAQ and ESR1 genes

3.4.2.1.2. A search for putative regulatory regions of the MDM4 gene identifies potential functional SNPs in the selected haplotype

3.4.2.1.3. Lack of evidence for regulatory function of regions in which potential functional PPP2R5E SNPs reside

3.4.2.1.4. Identification of candidate functional SNPs in putative regulatory regions of the CD44 gene

3.4.2.1.5. Search for functional SNPs in regulatory regions of the YWHAQ gene

3.4.2.1.6. Identification of candidate functional SNPs in introns 1 and 2 of the ESR1 gene

3.4.2.2. p53 network SNPs associate with allelic differences in cellular steady state mRNA levels

3.4.2.2.1. A selected haplotype in the MDM4 gene associates with reduced MDM4 mRNA expression in cells

3.4.2.2.2. A selected SNP in PPP2R5E associates with androgen receptor - dependent, allele-specific differences in mRNA levels in cells

3.4.2.2.3. Associations of the chemosensitivity SNPs in the CD44 and YWHAQ genes with their transcript levels in cells

3.4.2.2.4. An analysis of regulation of ESR1 mRNA expression in the NCI60 panel by gene amplification, promoter methylation and SNP rs2234693

3.4.3. Discussion

3.4.3.1. Potential functional SNPs upstream of the second MDM4 promoter might regulate MDM4 gene expression in an allele-specific manner

3.4.3.2. The expression of PPP2R5E mRNA is possibly regulated by a selected haplotype in the PPP2R5E gene through a yet unidentified SNP

3.4.3.3. Associations of chemosensitivity SNPs in the CD44 and YWHAQ genes with the gene expression in cells and identification of candidate functional SNPs

3.4.3.4. Identification of potential functional SNPs that might affect ESR1 mRNA expression

3.4.1. Introduction

In the previous chapters of this Thesis, we have incorporated characteristics of functional p53 pathway single nucleotide polymorphisms into three different screening approaches to identify other SNPs in this gene network that affect human cancer. The first approach evolved from the observation that the unusually long and frequent haplotype that harbors the G-allele of MDM2 SNP309 deviates significantly from the standard assumptions of selective neutrality, suggesting that the p53 pathway could be acted on by evolutionary selection pressure(s) (Chapter 3.1; (Atwal, Bond et al. 2007)). Similar signs of recent natural selection were subsequently observed by our group in the haplotype structure of the MDM4 gene, another key regulator of p53 (Atwal, Kirchhoff et al. 2009). Moreover, the neutral haplotypes of the MDM4 gene were shown to associate with a greater risk of developing breast and ovarian tumors in 5 different populations compared to those individuals with the selected haplotype (Atwal, Kirchhoff et al. 2009). Based on these studies, we reasoned that signatures of natural selection in other p53 network genes could help us identify p53 network SNPs that affect cancer. Indeed, we analyzed 142 p53 network genes, and found signs of recent natural selection in the PPP2R5E gene, a regulatory subunit of protein phosphatase 2A. Importantly, we observed that soft-tissue sarcoma patients with the selected haplotype associate with clinical differences compared to patients with the neutral haplotypes. For example, a selected SNP in PPP2R5E (ε-SNP2) associated with gender-dependent, significant allelic differences in the onset of (up-to 19.2 years; p=0.0002) and risk for STS (odds ratio, up-to 8.1; p=0.004), as well as overall survival (relative risk, up-to 2.68; p=0.045)

In the next screening approach, we searched for p53 network SNPs that associate with significant allelic differences in cellular chemosensitivities, a well-documented characteristic of both p53 codon72 and MDM2 SNP309 (Chapter 3.2). To do this, we utilized the NCI60 human tumor cell line anticancer drug screen to scan 142 p53 network genes and identified 7 SNPs in 6 genes that possess allelic differences in cellular growth responses to a large panel of cytotoxic chemotherapeutic agents. Cells with different genotypes at the identified loci associated with, on average, a different response for 79 of the 132 agents tested, whereby the homozygotes differed on average 4.9-fold in their respective drug sensitivities. We observed the greatest differences in SNPs in the YWHAQ (14-3-3tau, rs6734469, $p=5.6x10^{-47}$) and CD44 (rs187115, $p=8.1x10^{-24}$) genes. Importantly, sarcoma patients with the alleles of these SNPs that associated with weaker cellular growth responses to chemotherapeutics associated with poorer overall survival (up-to 2.89 relative risk, p=0.011) and an earlier age of diagnosis (up-to 10.7 years earlier, p=0.002).

Lastly, we searched for SNPs that, like MDM2 SNP309, could be also regulated by estrogen signaling and, therefore, associate with age- and gender-dependent differences in cancer risk (Chapter 3.3). To do this, we aimed to identify SNPs that associate with primarily female-specific differences in age-dependent incidence of STS like MDM2 SNP309. Using a tag-SNP approach and a medium throughput genotyping platform, we have genotyped 639 SNPs that tag 591 out of 766 (77%) known haplotypes of 113 p53 network genes in patients from the STS cohort. We found that the most significant allelic differences in the age of sarcoma diagnosis in females are found with two linked SNPs in the ESR1 gene (rs2234693 and rs9479130), whereby females, but not

males, carrying the major alleles of the SNPs show a 12 year earlier age of STS diagnosis (p=0.0374). Interestingly, the above-noted allelic differences in the age of STS onset for the ESR1 haplotype were restricted to females who carry the G-allele of MDM2 SNP309. Specifically, female MDM2 SNP309 G-allele carriers homozygous for the major alleles of the ESR1 polymorphisms were diagnosed on average 22.5 years (median 32.5 years) earlier than female G-allele carriers homozygous for the minor alleles. Interestingly, no significant allelic differences were observed for the ESR1 SNPs in females T/T in genotype for MDM2 SNP309. Importantly, we were able to confirm these associations between the genotypes at the ESR1 SNP and MDM2 SNP309 loci and the age- and gender-dependent cancer onset in two cancer patient cohorts. In sum, these data suggest that the ESR1 SNPs and MDM2 SNP309 can interact together in a gender-specific manner to affect cancer risk.

Taken together, the results of the three screening approaches suggest that allelic differences in signatures of natural selection, cellular chemosensitivities and gender- and age-dependent sarcoma incidence can be used to identify other p53 network genes that harbor potential functional SNPs that affect human cancer. However, the molecular and functional mechanisms underlying the above-described cellular and clinical associations of the SNPs in the PPP2R5E, YWHAQ, CD44, ESR1 and MDM4 genes identified in this Thesis and by our group, are as yet unknown and require further study.

The work described in this and the subsequent Chapter aims to begin to identify and characterize the causal SNP(s) in the haplotypes of the MDM4, PPP2R5E, YWHAQ, CD44 and ESR1 genes that associate with the clinical and cellular phenotypes. In this Chapter, we begin by describing the haplotypes associated with the identified SNPs, with

a focus on the known SNPs that are in close linkage with the identified polymorphisms. We proceed to identify those SNPs that reside in potential regulatory regions, which we define by a combination of evolutionary conservation, genic location, and/or experimental evidence. We identify multiple SNPs that could alter gene expression and go on to provide supporting evidence in both cell panels and heterozygous cells.

3.4.2. Results

3.4.2.1. The identification of potential causal SNPs that could underlie cellular and clinical phenotypes

3.4.2.1.1. Analysis of the linkage disequilibrium between polymorphisms in the MDM4, PPP2R5E, CD44, YWHAQ and ESR1 genes

We first developed a list of known SNPs that are in close linkage with the identified SNPs in the MDM4, PPP2R5E, YWHAQ, CD44 and ESR1 genes ((Atwal, Kirchhoff et al. 2009); Chapters 3.1 – 3.3) by using data from the International HapMap Project (HapMap Phase I, II and III; Chapter 1.3.3 (The_International_HapMap_Consortium 2005; Frazer, Ballinger et al. 2007; Altshuler, Gibbs et al. 2010)) and/or the 1000 Genomes Project Pilot 1 (Chapter 1.3.3; (Durbin, Abecasis et al. 2010)).

First, we have studied the evolutionary selected haplotype in the MDM4 gene, which has been recently identified by our group to associate with altered breast and ovarian cancer risk (Atwal, Kirchhoff et al. 2009). Using HapMap Phase I data and genotypes from a collection of 84 lymphoblastoid cell lines from the Coriell cell line

panel, this haplotype centered on SNP rs2369244 in the MDM4 gene has been previously shown to harbor 8 intragenic SNPs (Atwal, Kirchhoff et al. 2009). We have now extended this study utilizing HapMap Phase I and II data and found (within the intragenic region and 5kb upstream of the 5'UTR of the gene) that 14 known SNPs are significantly linked to SNP rs2369244 (r^2>0.8, CEU). Two SNPs reside in intron 1 of the MDM4 gene, eleven SNPs are scattered in introns 3-11 and one SNP in the 3'UTR region of the gene (Table 3.4.1; Figure 3.4.1). No other SNP(s) are significantly linked to any of those 14 SNPs (r^2>0.8) in the HapMap Phase I or II dataset. However, since this initial analysis, data from the HapMap Phase III and the 1000 Genomes Project have become available and suggest that 97 additional SNPs are significantly linked to SNP rs2369244 and the other 14 loci.

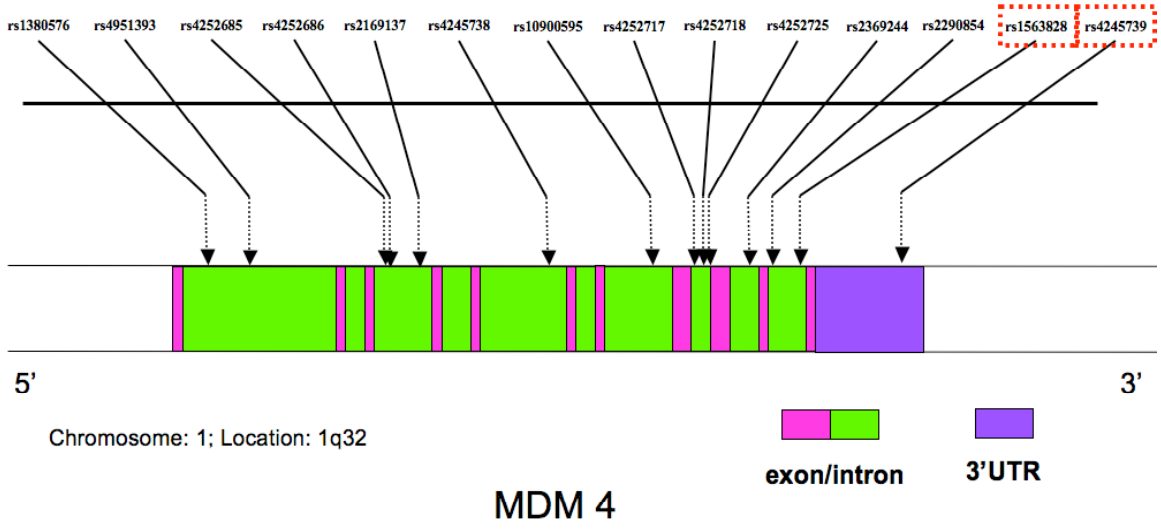


Figure 3.4.1. Haplotype structure and a schematic diagram of the MDM4 oncogene in individuals from the Centre d'Etude du Polymorphisme Humain collected in Utah, USA, with ancestry from northern and western Europe (CEU). The SNPs genotyped in this study are highlighted in a red box.

Table 3.4.1. List of SNPs from the haplotype centered on SNP rs2290854 and association with potential regulatory regions in the MDM4 gene

SNP	RSquared	DPrime	Conserved region (sequence alignment with mouse genome)	ENCODE Regulatory region (Functional element)	Genomic localization	Coordinate NCBI36/HG18 assembly, Chromosome 1
rs2290854	1	1	60%	H3K27ac in GM12878	Intron 10	202782648
rs4245738	1	1	60%	No	Intron 5	202772730
rs2369244	0.960	1	<50%	No	Intron 9	202781922
rs4252718	0.960	1	<50%	H3K4me1 in GM12878	Intron 8	202778818
rs4252686	0.960	1	60%	No	Intron 3	202763518
rs1563828	0.960	1	<50%	No	Intron 10	202783200
rs1380576	0.960	1	<50%	H3K4me1 in K562; Pol II ChIP in K562 and GM12878; H3K4me3 in K562 and GM12878; C-Myc ChIP in K562; Max ChIP in K562	Intron 1	202754901
rs4252685	0.960	1	60%	No	Intron 3	202763479
rs4252725	0.959	1	<50%	H3K4me1 in GM12878 and K562; H3K4me3 in GM12878	Intron 8	202779879
rs4252717	0.954	1	70%	H3K4me1 in GM12878	Intron 8	202778723
rs4951393	0.917	1	<50%	H3K4me1 in GM12878 and K562; H3K4me1 in GM12878 and K562; H3K27ac in GM12878	Intron 1	202756180
rs10900595	0.884	1	75%	No	Intron 7	202778225
rs4245739	0.847	1	70%	No	3' UTR	202785465
rs2169137	0.843	1	<50%	H3K4me1 in GM12878; Pol II in GM12878 and K562	Intron 3	202764536

We proceeded to study the loci in high linkage disequilibrium with the three linked SNPs in the PPP2R5E gene that we had shown to associate with an unusually long haplotype that might have undergone a recent selective sweep and allelic differences in

sarcoma onset and survival (Figure 3.1.1, Chapter 3.1.2.1). Two of the SNPs reside in intron 3 and one SNP in intron 2 of the PPP2R5E gene, which consists of 13 exons in total (Table 3.4.2; Figure 3.4.2). Utilizing data from all phases of the Hap Map Phase (I, II and III) and from Pilot 1 of the 1000 Genomes Project (r^2>0.8, CEU; search within 500kb of the tag SNPs), we find that these 3 SNPs are in near-perfect LD, and no other known SNP(s) are significantly linked to any of those 3 polymorphisms (Table 3.4.2).

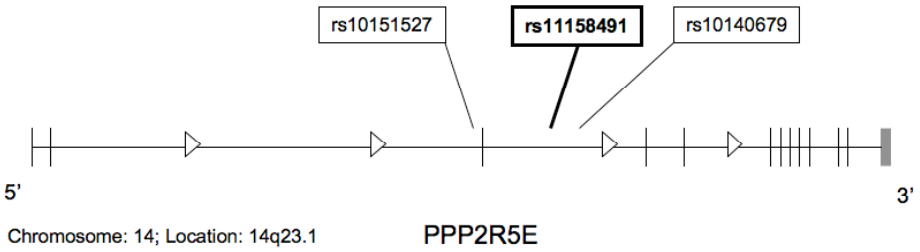


Figure 3.4.2. Schematic representation of the PPP2R5E gene and the approximate genic position of the identified linked SNPs. PPP2R5E ε-SNP2 in intron 3 of the PPP2R5E gene is depicted in bold. The exons are represented by vertical bars.

Table 3.4.2. List of SNPs from the haplotype centered on SNP rs11158491 and association with potential regulatory regions in the PPP2R5E gene

SNP	RSquared	DPrime	Conserved region (sequence alignment with mouse genome)	ENCODE Regulatory region (Functional element)	Genomic localization	Coordinate NCBI36/HG18 assembly, Chromosome 14
rs11158491	1	1	<50%	No	Intron 3	62980858
rs10140679	1	1	<50%	No	Intron 3	62971047
rs10151527	0.959	1	<50%	No	Intron 2	62992047

The CD44 SNP rs187115 (Chapter 3.2) that associates with the allelic differences in cellular chemosensitivities and sarcoma survival is in high linkage disequilibrium with 7 SNPs, all residing in intron 1 of the CD44 gene (Table 3.4.3; Figure 3.4.3A). No other

SNP(s) are significantly linked to any of those 7 polymorphisms (r^2>0.8, CEU; Hap Map I-III, 1000 Genomes Project, search within 500kb of the tag SNPs). The YWHAQ SNP rs6734469 that also associated with differences in cellular chemosensitivities and sarcoma survival (Chapter 3.2) is in high LD with 22 SNPs in introns 2,3 and 5 and within 5kb 3' of the YWHAQ gene (Table 3.4.4; Figure 3.4.3B). 10 of the 22 SNPs are significantly linked to 2 other SNPs in intron 2 (Table 3.4.4).

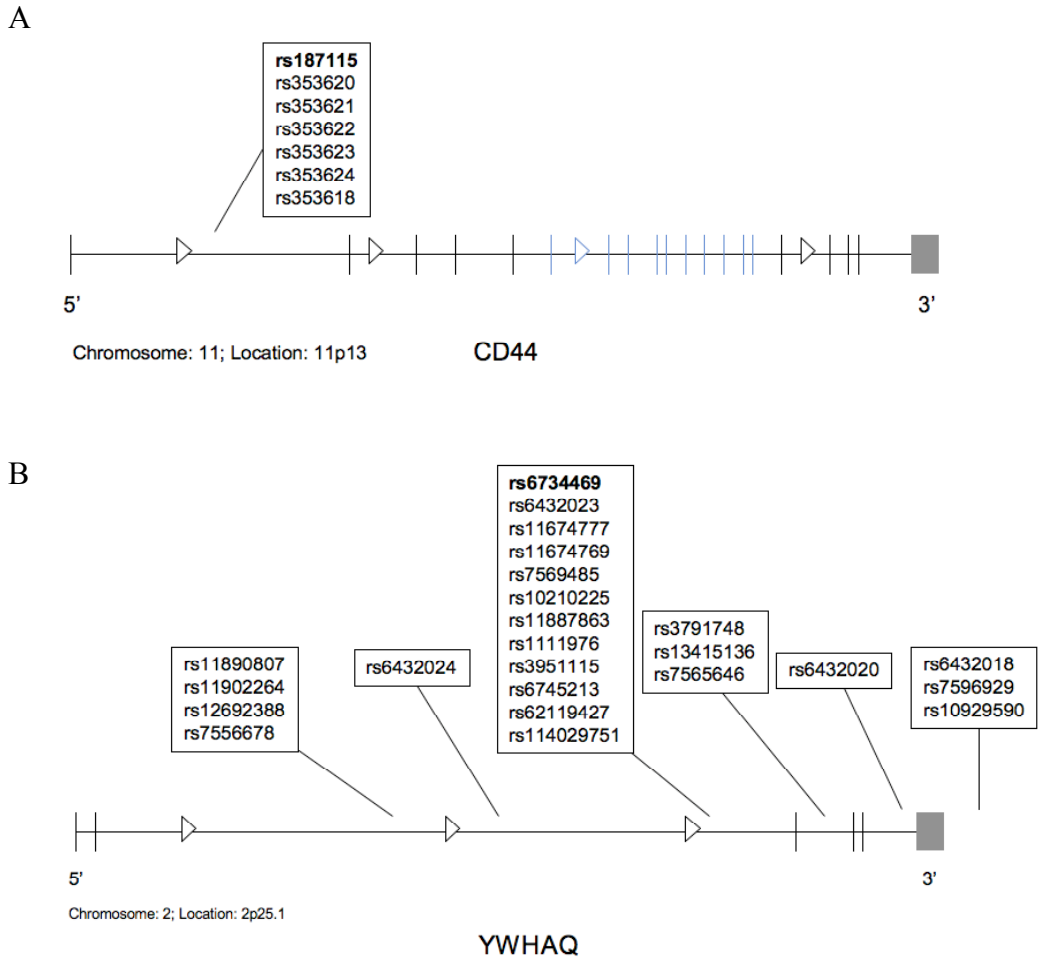


Figure 3.4.3. Schematic representation of the (A) CD44 and (B) YWHAQ genes and the approximate genic position of the identified linked SNPs. The alternatively spliced exons in the CD44 gene are highlighted in blue. The SNPs genotyped in this study are depicted in bold. The exons are represented by vertical bars.

Table 3.4.3. List of SNPs from the haplotype centered on SNP rs187115 and association with potential regulatory regions in the CD44 gene

SNP	RSquared	DPrime	Conserved region (sequence alignment with mouse genome)	ENCODE Regulatory region (Functional element)	Genomic localization	Coordinate NCBI36/HG18 assembly, Chromosome 11
rs187115	1	1	60%	No	Intron 1	35132735
rs353620	1	1	90%	No	Intron 1	35130712
rs353621	1	1	<50%	Pol II ChIP in GM12878	Intron 1	35129340
rs353622	0.897	1	<50%	Pol II ChIP in GM12878	Intron 1	35129233
rs353623	0.897	1	60%	Pol II ChIP in GM12878	Intron 1	35128698
rs353624	0.897	1	60%	Pol II ChIP in GM12878	Intron 1	35126431
rs353618	0.865	1	<50%	Pol II ChIP in GM12878	Intron 1	35131621

Table 3.4.4. List of SNPs from the haplotype centered on SNP rs6734469 and association with potential regulatory regions in the YWHAQ gene

SNP	RSquared	DPrime	Conserved region (sequence alignment with mouse genome)	ENCODE Regulatory region (Functional element)	Genomic localization	Coordinate NCBI36/HG18 assembly, Chromosome 2
rs6734469	1	1	<50%	H3K4me1 signal in HMEC and K562	Intron 2	9653351
rs6432023	1	1	<50%	H3K4me1 signal in K562	Intron 2	9654325
rs11674777	1	1	<50%	H3K4me1 signal in HMEC	Intron 2	9651801
rs11674769	1	1	65%	H3K4me1 signal in HMEC	Intron 2	9651729
rs3791748	0.964	1	<50%	No	Intron 3	9647266
rs13415136	0.964	1	<50%	No	Intron 3	9647159
rs7565646	0.964	1	<50%	No	Intron 3	9646330
rs11890807	0.964	1	<50%	H3K4me1 signal in HMEC	Intron 2	9671183
rs11902264	0.964	1	<50%	H3K4me1 signal in HMEC	Intron 2	9671537
rs12692388	0.964	1	<50%	No	Intron 2	9671920
rs7569485	0.927	0.963	<50%	No	Intron 2	9656225
rs10210225	0.927	0.963	<50%	No	Intron 2	9656676
rs11887863	0.927	0.963	<50%	No	Intron 2	9656880
rs1111976	0.927	0.963	<50%	No	Intron 2	9657006
rs3951115	0.927	0.963	<50%	No	Intron 2	9658012
rs6745213	0.927	0.963	<50%	No	Intron 2	9658997
rs6432024	0.927	0.963	<50%	no	Intron 2	9660585
rs6432018	0.927	0.963	<50%	Pol II ChIP in H1-hESC and K562	3' of gene	9639347
rs7596929	0.927	0.963	<50%	no	3' of gene	9635096
rs10929590	0.897	1	<50%	FOSL2 ChIP in HepG2; POL II ChIP in GM12878	3' of gene	9637251
rs7556678	0.897	1	<50%	no	Intron 2	9672817
rs6432020	0.856	0.925	<50%	no	Intron 5	9643371
rs62119427*	0.769	0.950	<50%	no	Intron 2	9659519
rs114029751*	0.741	0.958	<50%	H3K4me1 and H3K27ac signal in K562 and GM12878	Intron 2	9651487

*SNPs linked with Rsquare>0.8 to SNPs rs10210225, rs1111976, rs11887863, rs7596929, rs7569485, rs6745213, rs6432024, rs6432018, rs3951115. SNP rs62119427 is also linked to SNP rs7556678.

The ESR1 SNP^{rs2234693} that associates with allelic differences in sarcoma onset in a female-specific manner (Chapter 3.3) is in high linkage disequilibrium with one synonymous SNP (rs2077647) in consensus exon 1 (Kos, Reid et al. 2001) and a total of 15 SNPs in introns 1 and 2 of the ESR1 gene (Table 3.4.5; Figure 3.4.4). No other SNP(s) are significantly linked to any of those 17 polymorphisms (r^2>0.8).

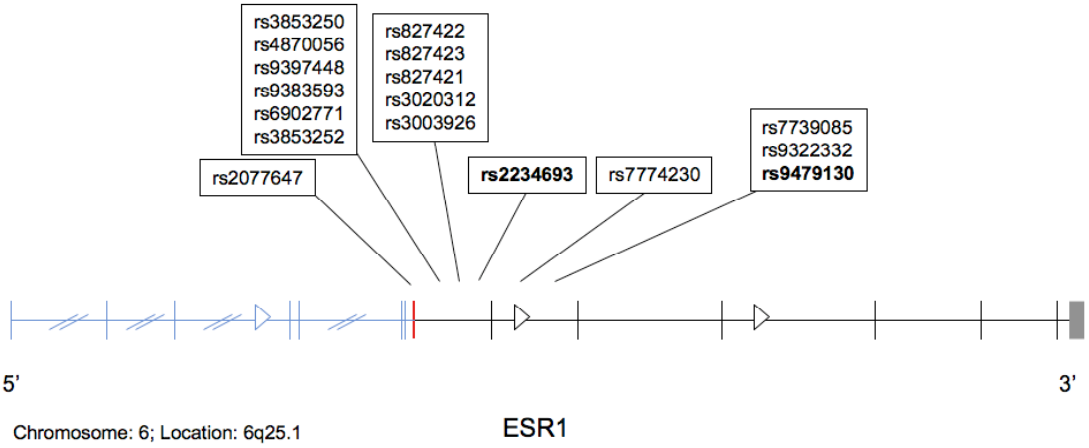


Figure 3.4.4. Schematic representation of the ESR1 gene and the approximate genic position of the identified linked SNPs. The alternatively spliced exons in the CD44 gene are highlighted in blue. The SNPs genotyped in this study are depicted in bold. The exons are represented by vertical bars. Consensus exon 1 is represented by the red vertical bar. The alternatively spliced exons upstream of exon 1 in the ESR1 gene are depicted in blue. The diagonal bars represent not shown large intronic regions.

Table 3.4.5. List of SNPs from the haplotype centered on SNP rs2234693 and association with potential regulatory regions in the ESR1 gene

SNP	RSquared	DPrime	Conserved region (sequence alignment with mouse genome)	ENCODE Regulatory region (Functional element)	Genomic localization	Coordinate NCBI36/HG18 assembly, Chromosome 6
rs2234693	1	1	75%	No	Intron 1	152205028
rs7774230	1	1	75%	No	Intron 2	152205932
rs7739085	1	1	<50%	No	Intron 2	152206241
rs9397448	1	1	<50%	No	Intron 1	152202759
rs3853250	1	1	<50%	No	Intron 1	152201593
rs9479130	1	1	<50%	No	Intron 2	152210149
rs3853252	1	1	<50%	No	Intron 1	152211940
rs9383593	0.965	1	<50%	No	Intron 1	152201319
rs6902771	0.933	1	<50%	No	Intron 1	152199574
rs9322332	0.932	1	<50%	No	Intron 2	152208494
rs4870056	0.899	1	<50%	No	Intron 1	152203920
rs827422	0.842	1	<50%	No	Intron 1	152198421
rs827423	0.842	1	<50%	No	Intron 1	152197890
rs827421	0.814	1	<50%	No	Intron 1	152198815
rs3020312	0.814	1	<50%	No	Intron 1	152197657
rs3003926	0.814	1	<50%	No	Intron 1	152197656
rs2077647	0.805	0.961	90%	H3K4me1 in HMEC and HUVEC; H3K27me3 in GM12878; DNase peak in KM-562	Exon 1	152170770

As seen in Tables 3.4.1 to 3.4.5, none of the SNPs in any of the genes result in changes in protein coding. Therefore, we make the assumption that any causal SNP would affect processes that would either directly or indirectly affect the cellular protein levels. We focus on processes that could be significantly affected by single base pair changes such as transcription, splicing and microRNA regulation. We first identify SNPs that are in known regulatory regions of the given gene, based on a survey of the literature. We then identify those SNPs that lie in a genic region that is often a site of cis-regulatory elements (such as proximity to an intron/exon boundary or in an UTR). As evolutionary conservation has been successfully used to identify both transcriptional and splicing regulatory elements (Pennacchio and Rubin 2001), we also identify those SNPs that

reside in conserved regions. Lastly, we identify the SNPs that associate with experimentally determined functional elements in the human genome in the Encyclopedia of DNA Elements (ENCODE) Project (http://genome.ucsc.edu/ENCODE/) (Birney, Stamatoyannopoulos et al. 2007).

The ENCODE Consortium is an international collaboration of research groups funded by the National Human Genome Research Institute (NHGRI) (Birney, Stamatoyannopoulos et al. 2007). Its aim is to build a comprehensive parts list of functional elements in the human genome, including those that act at the protein and RNA levels, and regulatory elements that control cells and circumstances in which a gene is active (Birney, Stamatoyannopoulos et al. 2007). The ENCODE project covers a portion of the genome equal to about 1% (30Mb) (Birney, Stamatoyannopoulos et al. 2007). 50% of the sample area selected for study under this phase was manually selected based on the presence of well-studied genes and the availability of comparative data, whilst the other 50% was selected at random (Birney, Stamatoyannopoulos et al. 2007). Specifically, the experimental techniques utilized in the ENCODE Project aim to study chromatin accessibility and histone modification patterns, sequence-specific factors, replication, comparative sequence analysis, transcription, polymorphisms and computational analysis (Birney, Stamatoyannopoulos et al. 2007). It utilizes 8 common cell types, GM12878 (lymphobalstoid), H1-hESC (human embryonic stem cells), K562 and K562b (leukemia), HeLa-S3 (cervical carcinoma), HepG2 and HepG2b (liver carcinoma) and HUVEC (Human Umbilical Vein Endothelial Cell) as well as up to 194 additional cell lines for selected experiments (http://genome.ucsc.edu/ENCODE/cellTypes.html) (Birney, Stamatoyannopoulos et al.

2007). Given the data available at the time of writing of this Thesis, we have chosen to focus on experimental data aiming to study chromatin accessibility (DNase I sensitivity), histone modification patterns (histone methylation signatures: H3K4me1, H3K4me3, H3K27me3, H3K27ac) and sequence-specific factors (18 sequence-specific transcription factors such as Myc, Max, YY1 and components of the general transcription machinery, such as RNA polymerase II, TAF1 and TFIIB/GTF2B) (Birney, Stamatoyannopoulos et al. 2007) as those might denote putative transcriptional enhancer or promoter regions (Birney, Stamatoyannopoulos et al. 2007; Koch, Andrews et al. 2007; Rada-Iglesias, Bajpai et al. 2011)).

3.4.2.1.2. A search for putative regulatory regions of the MDM4 gene identifies potential functional SNPs in the selected haplotype

In the above-described analysis (Section 3.4.2.1.1), we have identified that the selected haplotype in the MDM4 gene harbors a total of 14 significantly linked intragenic SNPs (Table 3.4.1, Figure 3.4.1). The two polymorphisms in intron 1 (rs1380576, rs4951393) reside in a characterized regulatory region of the MDM4 gene, located approximately 5kb 5' of a second MDM4 promoter (Phillips, Teunisse et al. 2010). In line with the region's characterized regulatory function is the observation that both SNPs, and particularly SNP rs1380576, associate with the expression of experimentally determined functional elements in the ENCODE project (Table 3.4.1). Specifically, the region surrounding SNP rs1380576 harbors an H3K4me1 signature (mono-methylation of Lysine 4 in histone H3) in K562 cells and an H3K4me3 signature (tri-methylation of

Lysine 4 in histone H3) in K562 and GM12878 cells. Both histone methylation patterns are indicative of actively transcribed regulatory regions (Koch, Andrews et al. 2007; Rada-Iglesias, Bajpai et al. 2011). Furthermore, binding of RNA Polymerase II (Pol II) to the region surrounding SNP rs1380576 has been shown in a ChIP experiment in K562 and GM12878 cells, and, in K562 cells, a binding of the transcription factors c-Myc and Max has also been demonstrated. The binding of these transcription factors and of Pol II is a common occurrence near transcriptional start sites (TSS) and in genomic regions enriched for regulatory elements (Birney, Stamatoyannopoulos et al. 2007). Histone modification patterns indicative of active transcription and open chromatin conformation are also reported for the second SNP in intron 1 (rs4951393; Table 3.4.1). Besides an H3K4me1 signature, it also harbors an H3K27ac modification (acetylation of histone H3 on Lysine 27). Lysine acetylation on histone H3 is generally associated with activation of transcription and an open chromatin formation (Koch, Andrews et al. 2007).

An additional SNP that could potentially affect the gene expression of MDM4 is SNP rs4245739. This SNP is located at the margins of a well-conserved region in the 3'UTR (Table 3.4.1). In fact, a recent report suggests that this SNP affects the binding of the microRNA hsa-miR-191, thereby affecting its regulation of MDM4 (Wynendaele, Bohnke et al. 2010). The remaining 11 SNPs reside in introns 3-10 and show weaker signs of potential regulatory regions (Figure 3.4.1, Table 3.4.1). For instance, SNP rs10900595 is located in a well-conserved region (about 75% sequence alignment with the mouse genome) in exon 7, which could constitute a distal enhancer, although no additional signs in the ENCODE data were noted. Taken together, these data suggest that polymorphisms from the selected MDM4 haplotype, particularly both SNPs in intron 1

and a SNP in the 3'UTR, reside in a region of the MDM4 gene possibly involved in the regulation of transcription or mRNA processing, respectively, and could therefore affect the activity of the MDM4 gene by regulating its expression on the mRNA level.

3.4.2.1.3. Lack of evidence for regulatory function of regions in which potential functional PPP2R5E SNPs reside

In the PPP2R5E gene, none of the 3 identified SNPs in the selected haplotype (Section 3.4.2.1.1; Table 3.4.2, Figure 3.4.2) is located in a known regulatory region of the gene. Furthermore, the regions in which they reside do not show signs of conservation between species or evidence in the ENCODE data set that they could be functional regulatory elements (Table 3.4.2).

3.4.2.1.4. Identification of candidate functional SNPs in putative regulatory regions of the CD44 gene

Neither of the 7 identified SNPs in intron 1 the CD44 gene is located in a known regulatory genomic region (Section 3.4.2.1.1; Table 3.4.3, Figure 3.4.3A). The CD44 gene contains 20 exons, and its function is thought to be, in part, regulated by alternative splicing of the gene (Ponta, Sherman et al. 2003; Liu and Jiang 2006). The first and the last 5 exons are transcribed in all of the 6 known CD44 isoforms and the remaining 10 exons are alternatively spliced in a tissue-specific manner (Figure 3.4.3A; (Ponta, Sherman et al. 2003; Liu and Jiang 2006). Of the 7 SNPs in intron 1 of the CD44 gene,

one (rs353620) resides in a strongly conserved region (up-to 90% sequence alignment with the mouse genome), and 5 other SNPs show some level of evidence of residing in a regulatory region in the ENCODE dataset (Table 3.4.3). Specifically, the binding of Pol II (as determined by ChIP) to those regions has been reported by the ENCODE Project in GM12878 cells, indicating an association with potential TSSs. Taken together, the analysis of potential regulatory regions in the linked CD44 SNPs provides some evidence that one or more of the identified SNPs might be functional polymorphisms and regulate the transcription of the CD44 gene.

3.4.2.1.5. Search for functional SNPs in regulatory regions of the YWHAQ gene

None of the 24 identified SNPs in the YWHAQ gene is located in a genomic region known to harbor regulatory elements (Section 3.4.2.1.1; Table 3.4.4, Figure 3.4.3B). Only one of the 24 linked SNPs in the YWHAQ gene (intron 2, rs11674769) resides in a moderately conserved region, while 7 SNPs (including rs11674769) in intron 2 are located in regions with some evidence for regulatory function in the ENCODE data set (Figure 3.4.3B, Table 3.4.4). Specifically, all 7 SNPs associate with regions with an H3K4me1 signature in HMEC and/or K562 cell lines, indicative of actively transcribed regulatory regions near TSSs (Koch, Andrews et al. 2007). In addition, SNP rs114029751, which is closely linked (r^2>0.8) to 10 of the 24 SNPs that comprise the identified YWHAQ haplotype associates with the margins of a putative regulatory region in intron 2, which harbors an H3K27ac signature, a characteristic of an open chromatin formation (Table 3.4.4). Taken together, the analysis of potential regulatory regions

harboring the linked YWHAQ SNPs provides some support that one or more of the identified SNPs could affect the mRNA levels of the YWHAQ gene.

3.4.2.1.6. Identification of candidate functional SNPs in introns 1 and 2 of the ESR1 gene

The ESR1 gene contains 8 translated exons with the transcriptional start site located in consensus exon 1 (Kos, Reid et al. 2001). Several exons upstream of exon 1 have been reported in the literature, which are predominantly spliced to exon 1 and termed 1', B, C, D, E1, E2, F, Ha, Hb, T1 and T2 (Figure 3.4.4; (Kos, Reid et al. 2001)). Their expression has been suggested to depend on the developmental stage and tissue type (Kos et al 2001). Alternatively, it has been shown in MCF7 cells that exon F can be spliced directly to exon 2 in approximately 10% of transcripts originating form the F-promoter (Flouriot et al 2000). Regulation of such splicing could be affected by polymorphic variation within exon 1. However, the synonymous SNP rs2077647 is located 98bp from the nearest exon/intron boundary and therefore rather unlikely to affect splicing, as exonic regulatory regions such as exonic splicing enhancers (ESE) or exonic splicing silencers (ESS) are thought to preferably reside within approximately 25bp and 60bp of (putative) exon/intron boundaries, respectively (Blencowe 2000; Chew, Baginsky et al. 2000; Fairbrother, Yeh et al. 2002; Fairbrother, Yeo et al. 2004; Wang, Rolish et al. 2004; Szeszel-Fedorowicz, Talukdar et al. 2006; Wang and Burge 2008; Wang, Crystal et al. 2009; Raponi, Kralovicova et al. 2011). Furthermore, although SNP rs2077647 resides in a highly conserved exonic region, it is itself not conserved between species, suggesting that it lacks a regulatory function. However, a recent report suggests

that synonymous SNPs residing in coding regions might regulate the expression of genes in which they reside by alteration of microRNA binding (Brest, Lapaquette et al. 2011). We therefore perform an in silica analysis using publicly available algorithms (SnipMir, RegRNA and Patrocles; Materials and Methods; Chapter 2.3.7.) to predict whether SNP rs2077647 might affect microRNA binding. Indeed, the C-allele of the SNP is predicted to change the binding affinity to the microRNA hsa-miR-1296, and introduce a binding site for microRNA hsa-miR-105*, while the T-allele introduces a potential binding site for the microRNA hsa-miR-222* (Figure 3.4.5). These changes in microRNA binding affinities have therefore the potential to result in allelic differences in ESR1 mRNA expression.

Figure 3.4.5. ESR1 SNP rs2077647. *In silico* prediction of microRNA and ESR1 mRNA interactions showed differences in binding of the microRNAs hsa-miR-1296, hsa-miR-222* and hsa-miR-105*. The C-allele changes the predicted binding pattern of hsa-miR-1296, while hsa-miR-222* and hsa-miR-105* are predicted to introduce new microRNA binding sites for the T- and C-alleles, respectively. Vertical line and colon indicate agreement with, or deviation from, consensus sequence alignment, respectively. SNP rs2077647 is depicted in red.

We next analyzed whether any of the linked intronic SNPs resides in a putative regulatory region of the ESR1 gene (Table 3.4.5), as neither intron 1 nor 2 have been reported in the literature to date to harbor any regulatory elements. Two SNPs (rs2234693 in intron 1 and rs7774230 in intron 2; located 396bp and 317bp from nearest exon/intron boundary, respectively) reside in well-conserved regions flanking exon 2, indicative of a potential role in gene regulation (Table 3.4.5). The ENCODE dataset suggests that the ESR1 gene is silenced in at least two of the utilized common cell lines (GM12878 and K562, a leukemia cell line with known methylation of the ESR1 promoter (Issa, Zehnbauer et al. 1996), for which ENCODE provides the most comprehensive information on functional elements. Specifically, the known promoter region of the ESR1 gene expresses an H3K27me3 signature (indicates transcriptionally inactive heterochromatin (Koch, Andrews et al. 2007)) and the lack of an H3K4me1 or H3K4me3 signature. Furthermore, the H3K4me1 and H3K4me3 signatures are only weakly expressed in the known ESR1 promoter region in the other common cell lines, suggesting that data obtained from this database might not be suitable enough to search for putative regulatory regions within the ESR1 gene. We find no evidence for regulatory function in the regions surrounding the 17 SNPs from the identified ESR1 haplotype in the ENCODE database (Table 3.4.5). Taken together, despite the apparent limited benefit of the ENCODE dataset in the study of ESR1 regulation, the analysis of sequence conservation suggests that this haplotype might harbor a polymorphism in a potential distal enhancer intronic region flanking exon 2 (such as SNPs rs2234693 and rs7774230), that could affect the expression of the ESR1 gene.

3.4.2.2. p53 network SNPs associate with allelic differences in cellular steady state mRNA levels

3.4.2.2.1. A selected haplotype in the MDM4 gene associates with reduced MDM4 mRNA expression in cells

In the previous sections, we have identified potential candidate SNPs that could affect the expression of the MDM4, CD44, YWHAQ and ESR1 genes and therefore underlie the allelic differences seen in the cellular and clinical phenotypes described in the previous chapters. To test this possibility, we will begin by exploring allelic differences of these loci in the steady state mRNA levels of the respective genes in the NCI60 cell line panel. We first explore potential allelic differences in MDM4 mRNA levels of a MDM4 SNP that differentiates the selected haplotype from the neutral MDM4 haplotypes and associates with differences in cancer risk (Atwal, Kirchhoff et al. 2009). To do this, we genotype the 59 cell-lines that make up the NCI60 cell panel for the MDM4 SNP (rs1563828 T/C, Figure 3.4.1), whereby the C-allele associates with the selected MDM4 haplotype. We measure the MDM4 mRNA levels for each cell line using real-time quantitative PCR with a commercially available TaqMan Gene Expression Assay that utilizes a probe that covers exons 3 and 4 and therefore assess the full-length MDM4 isoform and the variants MDM4-A, MDM4-XALT1 and MDM4-S (Mancini et al 2009). All RNAs were isolated from logarithmically growing cells and purified by the Developmental Therapeutics Program (DTP) of the National Cancer Institute (NCI) in Bethesda, MD, USA. The MDM4 measurements are normalized to the transcript levels of glyceraldehyde 3-phosphate dehydrogenase (GAPDH) and the values are reported using

the delta Cycle-threshold (ΔCt) method. Hence, lower numerical values correspond to higher transcript levels and a ΔCt difference of 1 corresponds to approximately a 2-fold difference in relative levels.

We identify 8 cells in the NCI60 cell panel that are homozygous for the T-allele, 25 that are heterozygous and 26 that are C/C in genotype. Interestingly, the cell lines T/T in genotype express significantly, 2.6-fold higher median MDM4 levels than those C/C in genotype, while those heterozygous at this locus express intermediate levels ($p=0.031$; Jonckheere-Terpstra Test; Figure 3.4.6). Specifically, the median ΔCt for cells homozygous for the T-allele is 7.28, whereas heterozygous cells and cells homozygous for the C-allele express median MDM4 mRNA levels of 8.43 and 8.66 ΔCt, respectively (Fig 3.4.6). These results lend support to the hypothesis that the identified potential candidate SNPs could affect MDM4 gene expression (Section 3.4.2.1) and underlie the allelic differences seen in breast and ovarian cancer risk (Atwal, Kirchhoff et al. 2009).

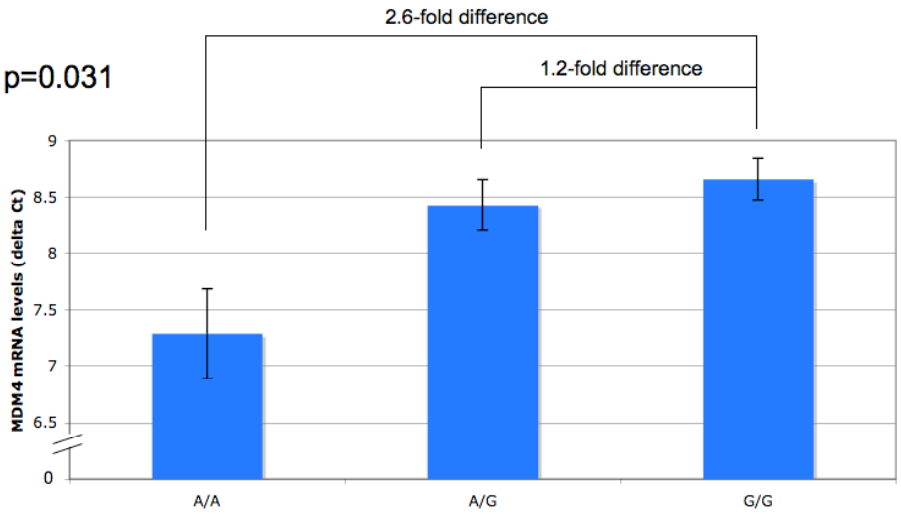


Figure 3.4.6. A selected haplotype in the MDM4 gene centered on SNP rs1563828 associates with decreased MDM4 transcript levels in untreated cells from the NCI60 panel. Cell lines homozygous for the minor ancestral A-allele of SNP rs1563828 express significantly 2.6-fold higher MDM4 mRNA levels than those G/G in genotype, while cells heterozygous at this locus are intermediate. The p-values are calculated using a Jonckheere-Terpstra Test. The values on the Y-axis represent ΔCt readings, whereby low numerical numbers correspond to high relative mRNA levels. The inferred fold-difference in the mRNA expression between the groups/genotypes is depicted above the bars.

3.4.2.2.2. A selected SNP in PPP2R5E associates with androgen receptor-dependent, allele-specific differences in mRNA levels in cells

The first two introns of various genes have been shown to harbor enhancers that are important in the regulation of upstream promoter activity and the expression of the respective genes (Bornstein, McKay et al. 1988; Rippe, Lorenzen et al. 1989; Alder, Yoshinouchi et al. 1992; MacKenzie and Quinn 1999; Bharadwaj, Trainor et al. 2003; Klenova, Scott et al. 2004). The analysis of SNPs, which are in high linkage disequilibrium with the selected PPP2R5E haplotype identified one SNP (rs10151527) in

intron 2 of the PPP2R5E gene that could potentially reside in an enhancer region not identified in the ENCODE Project and affect the expression of PPP2R5E. Therefore, in order to begin to explore the possibility that the allelic differences noted for the selected SNP in PPP2R5E, ε-SNP2, in sarcoma risk and survival could be explained by differences in expression levels, we determine the SNP genotypes in the NCI60 cell panel. We measure the steady state PPP2R5E mRNA levels for each cell line using RT-PCR and report the values using the ΔCt method. Ten cells are homozygous for the T-allele, 15 are heterozygous and 34 are C/C in genotype. Interestingly, cell lines harboring at least one T-allele express significantly, on average 1.41-fold higher PPP2R5E levels than those C/C in genotype ($p=0.027$; Mann-Whitney Test, Figure 3.4.7A). Specifically, cells homozygous for the T-allele (median ΔCt 7.46) express on average 1.47-fold higher PPP2R5E levels than those C/C in genotype (median 8.02 ΔCt), while those heterozygous at this locus are intermediate (median ΔCt 7.56; $p=0.048$; Jonckheere-Terpstra Test; Figure 3.4.7A). These results support the hypothesis that SNP rs10151527 in intron 2 of the gene, or another closely linked SNP(s), affects the cellular PPP2R5E levels by regulating its mRNA levels, possibly through transcriptional regulation.

The reported clinical associations of the PPP2R5E ε-SNP2 with cancer risk in the sarcoma cohort are most pronounced in the male patients (Chapter 3.1.2.1). One potential explanation for this gender-difference could be a differential response of this locus, or associated loci, to gender-specific hormonal signaling. To begin to explore this possibility, we measure and characterize the steady state androgen receptor (AR) and estrogen receptor 1 (ESR1) mRNA levels in the NCI60 cell panel. Importantly, we

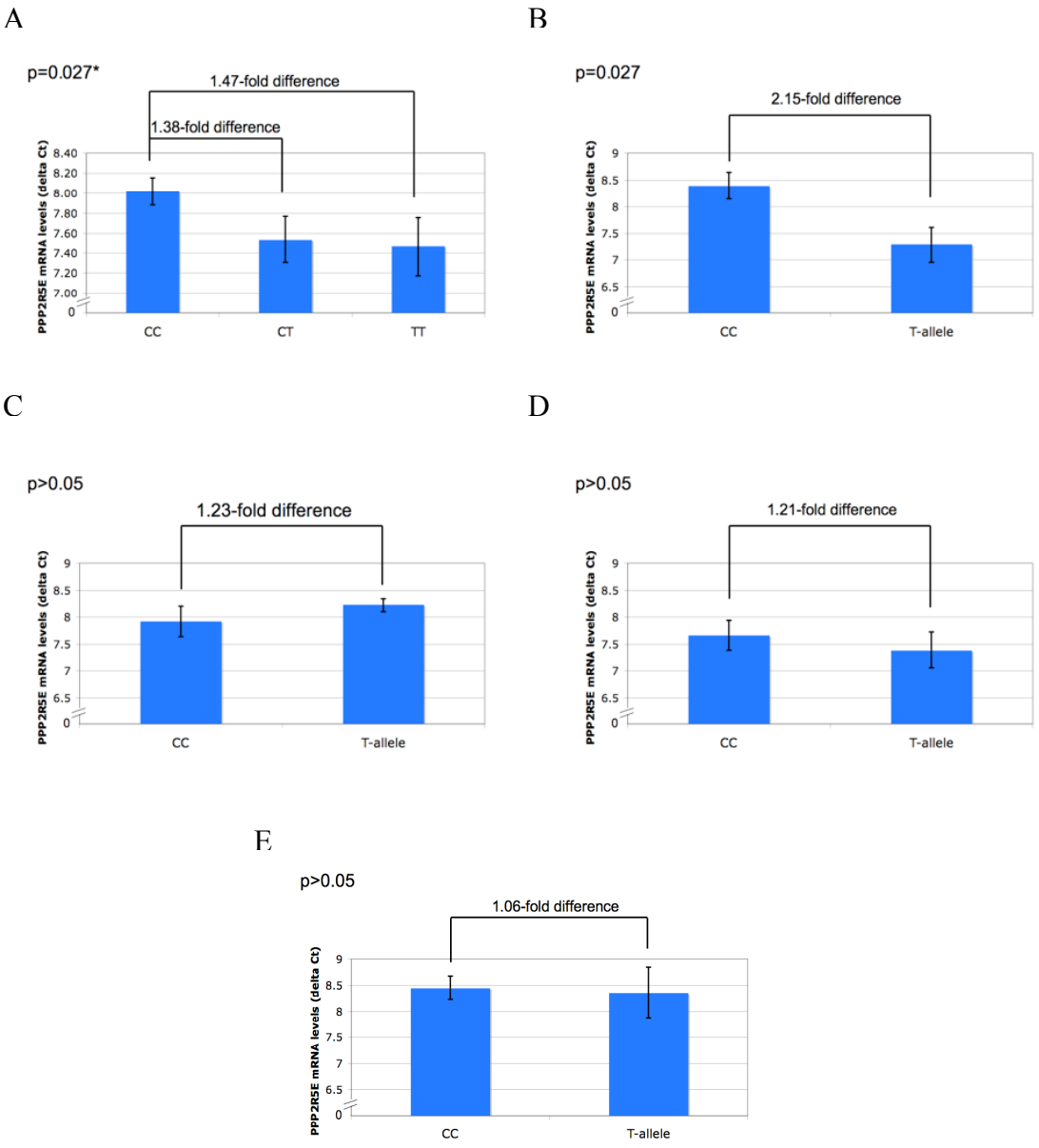


Figure 3.4.7. The selected haplotype in the PPP2R5E gene associates with elevated PPP2R5E steady-state transcript levels in the NCI60 panel. (A) Cell lines T/T in genotype express significantly, on average 1.47-fold higher MDM4 levels than those C/C in genotype, while those heterozygous at this locus are intermediate. (B) In cells that express very high levels of AR mRNA, those that harbor the T-allele of ε-SNP2 show 2.15-fold higher PPP2R5E mRNA levels than cells C/C in genotype. (C) In contrast, no significant allelic differences are observed in cells, which express very low levels of AR. The allelic differences at the PPP2R5E ε-SNP2 locus are equally pronounced in cells expressing (D) very high and (E) very low ESR1 mRNA levels. *All p-values are calculated using a Mann-Whitney Test comparing the expression in cells carrying the minor T-allele to those C/C in genotype.

observe that the two prostate cancer cell lines included in the NCI60 panel, PC-3 and DU-145, which are known to be AR negative (Tilley, Wilson et al. 1990; Alimirah, Chen et al. 2006), are found in the group of NCI60 cells that express significantly lower AR mRNA levels (Chapter 2.1.3.2, Table 2.8). Furthermore, we observe that both cell lines included in the NCI60 with well-documented high ERα expression, the breast cancer cell lines MCF7 and T47D, have the highest ESR1 mRNA levels of the entire NCI60 panel. In contrast, the breast cancer cell lines MDA-MB 435, HS578, MDA-MB-231 and BT-549, which are reported to be ERα negative, express significantly lower levels (Thompson, Paik et al. 1992; Lu, Mira-y-Lopez et al. 2005; Neve, Chin et al. 2006); Chapter 2.1.3.2, Table 2.8).

As described above, in all the cell lines, the cells harboring a T-allele have on average 1.41-fold higher PPP2R5E mRNA levels ($p=0.027$; Mann-Whitney Test, Figure 3.4.7A). Interestingly, when we divide the NCI60 panel into four equal groups according to the AR mRNA expression [those with very high ($n=14$, <10.0 ΔCt), high ($n=15$, <14.0 ΔCt), low ($n=15$, <21.0 ΔCt) and very low transcript levels ($n=15$, ≥ 21.0 ΔCt)], the allelic differences in PPP2R5E expression for PPP2R5E ε-SNP2 are particularly pronounced in the very high AR expressing cell lines. Specifically, in cells that express very high levels of AR mRNA, those cells that harbor the T-allele of ε-SNP2 ($n=7$) show 2.15-fold higher PPP2R5E mRNA levels than cells C/C in genotype ($n=7$, $p=0.027$; Mann-Whitney Test, Figure 3.4.7B). In contrast, in cells that express very low levels of AR, no significant allelic differences between the genotypes are observed, whereby cells with a C/C genotype ($n=11$) express 1.23-fold higher PPP2R5E mRNA levels than those harboring the T-allele ($n=4$, $p>0.05$, Mann-Whitney Test, Figure 3.4.7C). In contrast to the

observation made with the androgen receptor, when we divide the NCI60 panel into four equal groups according to the ESR1 mRNA expression [in those with very high (n=14, ≤15.4 ΔCt), high (n=15, ≤19.1 ΔCt), low (n=15, ≤22.4 ΔCt) and very low (n=15, >22.4) transcript levels], the allelic differences for PPP2R5E ε-SNP2 in PPP2R5E expression are neither pronounced nor significant in the very high ESR1 expressing cell lines. Specifically, in cell lines that express very high levels of ESR1 mRNA, those cells that harbor the T-allele of ε-SNP2 (n=9) show 1.21-fold higher PPP2R5E mRNA levels than cells C/C in genotype (n=5; p=0.364; Mann-Whitney Test, Figure 3.4.7D). In comparison, in cells, which express very low levels of ESR1, cells harboring the T-allele (n=3) express 1.06-fold higher PPP2R5E mRNA levels than those C/C in genotype (n=12; p=0.734, Mann-Whitney Test, Figure 3.4.7E). Together, these associations suggest that the primarily male-specific associations of the PPP2R5E ε-SNP2 with sarcoma risk could be a differential response of this locus, or associated loci, to androgen, but not estrogen, signaling, resulting in allelic differences in PPP2R5E mRNA levels.

3.4.2.2.3. Associations of the chemosensitivity SNPs in the CD44 and YWHAQ genes with their transcript levels in cells

We next explore potential associations of the SNPs in the CD44 and YWHAQ genes with allelic differences of cellular steady state mRNA levels. We measure the steady state CD44 and YWHAQ mRNA levels for each cell line as described above and explore potential associations between the genotypes of each SNP and the respective transcript levels. Interestingly, we find that the minor, C-allele of the CD44 SNP that

requires higher drug concentrations to suppress tumor growth in the NCI60 panel, associates with significantly 4.35-fold higher mRNA transcript levels than cells homozygous for the T-allele (p=0.021, Mann-Whitney Test, Figure 3.4.8). Specifically, cells C/C (6.52 ΔCt) and C/T (5.49 ΔCt) in genotype express on average 2.79-fold and 5.68–fold more CD44 mRNA than cells homozygous for the T-allele (8.00 ΔCt), respectively (Figure 3.4.8). Together, these data support the hypothesis that the CD44 SNP, or associated loci, could affect the activity of the CD44 gene by affecting its mRNA steady state levels.

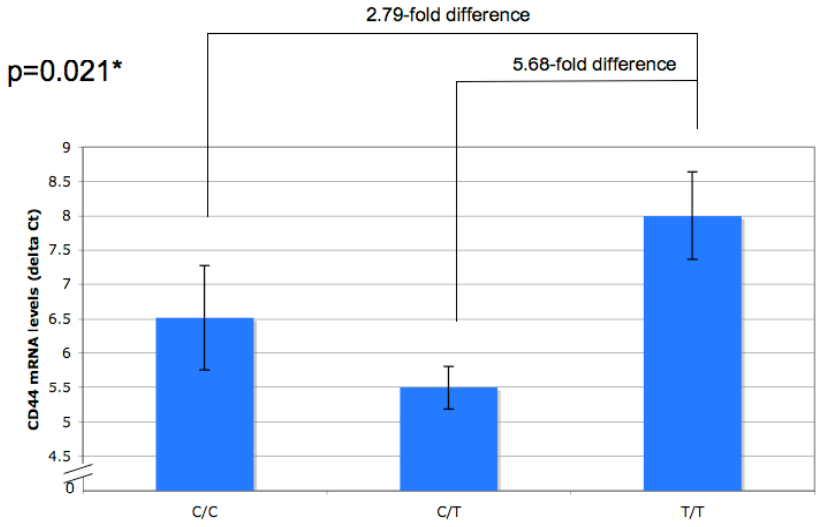


Figure 3.4.8. The minor allele of CD44 SNP, which requires higher drug concentrations to suppress tumor growth in the NCI60 panel, associates with significantly, higher mRNA transcript levels than cells homozygous for the major T-allele. The p-values are calculated using a Mann-Whitney Test.

A similar, non-significant, trend is observed for the YWHAQ SNP, whereby the A-allele that requires higher drug concentrations to suppress tumor growth in the NCI60 panel, also associates with higher YWHAQ transcript levels compared to the G-allele.

Specifically, cells A/A (median 4.93 ΔCt) and A/C (median 5.29 ΔCt) in genotype express 1.48-fold and 1.15-fold more mRNA, respectively, than cells homozygous for the C-allele (median 5.50 ΔCt) at the YWHAQ locus (p=0.088, Jonckheere-Terpstra Test, Figure 3.4.9). When we stratify the cells according to the p53 mutational status, as the allelic differences in chemosensitivity profiles are only observed in p53 wild-type (n=18), but not p53 mutant cells (n=41), no significant allelic differences in YWHAQ mRNA expression are observed in either group (data not shown).

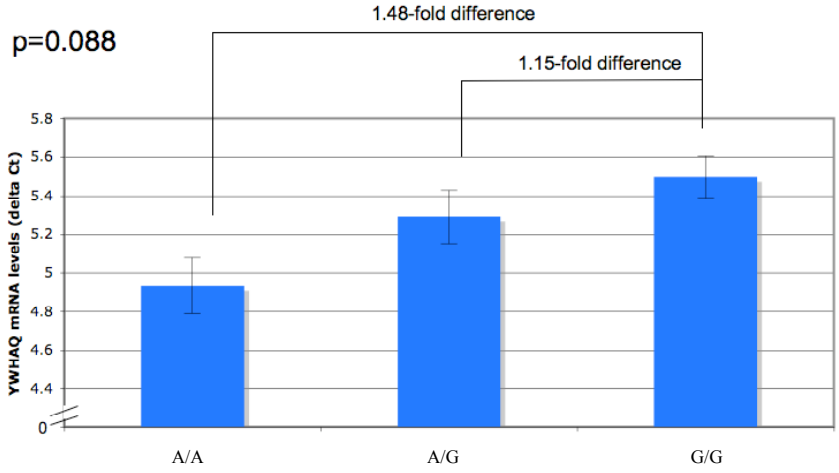


Figure 3.4.9. An insignificant trend is observed for the YWHAQ SNP, whereby cells homozygous for the A-allele, which require higher drug concentrations to suppress tumor growth in the NCI60 panel, associate with higher YWHAQ transcript levels compared cells G/G in genotype. The p-values are calculated using a Jonckheere-Terpstra Test

3.4.2.2.4. An analysis of regulation of ESR1 mRNA expression in the NCI60 panel by gene amplification, promoter methylation and SNP rs2234693.

The allelic differences in age- and gender-dependent onset and risk of cancer development described in Chapter 3.3 lend strong support to the hypothesis that SNPs linked to rs2234693 could ultimately alter the activity of the estrogen receptor in cells. Indeed, studies with human brain tissue of schizophrenic patients support this hypothesis, as allelic differences of these SNPs were noted in estrogen receptor mRNA levels (Weickert, Miranda-Angulo et al. 2008). Specifically, brain tissues obtained from individuals carrying the T-allele at the ESR1 SNP rs2234693 had significantly higher ESR1 mRNA expression than those with a C/C genotype (Weickert, Miranda-Angulo et al. 2008). In our study, the T-allele associates with an earlier age of soft-tissue sarcoma in female patients. To test the hypothesis that the observed clinical differences are indeed due to allelic differences in ESR1 mRNA transcription, we measure the ESR1 transcript levels in the NCI60 human tumor cell line panel as described above. However, testing of this hypothesis in the NCI60 panel is complicated by the high frequency of ESR1 silencing in cancer and cultured cells (Widschwendter, Siegmund et al. 2004; Holst, Stahl et al. 2007).

The expression of estrogen receptor is regulated by both genetic and epigenetic mechanisms (Darbre and King 1988; Widschwendter, Siegmund et al. 2004; Holst, Stahl et al. 2007; Ramos, Camargo et al. 2010). Specifically, in approximately 21% of breast tumors, the ESR1 gene is strongly overexpressed due to gene amplification (Holst, Stahl et al. 2007), and is silenced in approximately 41% of breast cancers by promoter

methylation (Ramos, Camargo et al. 2010). To asses the gene copy number of ESR1 and its methylation status in the NCI60 cell panel, we determine the ESR1 gene copy number (CN) by an RT-PCR based technique using an Applied Biosystems' TaqMan Copy Number Assay and incorporate the methylation status of the ESR1 promoter as reported by Shen et al (Shen, Kondo et al. 2007). 23 cell lines (39%) have an amplified ESR1 gene (CN≥2), 26 (44%) a normal copy number (CN=2) and 10 cell lines (17%) have a deletion of one ESR1 allele (CN=1; Table 3.4.6). ESR1 gene promoter methylation is

Table 3.4.6. Gene copy number and promoter methylation status of the ESR1 gene in the NCI60 panel

ESR1 gene	n (%)
Copy number	
≥ 2	23 (39%)
2	26 (44%)
1	10 (17%)
Promoter methylation*	
Significant	33 (56%)
None	24 (41%)

*The promoter methylation of 2 cell lines is not known

observed in 33 cell lines (56%), 24 cell lines (41%) show no significant promoter methylation, while no data on the methylation status of 2 cell lines (3%) is available (Table 3.4.6). As could be expected, both ESR1 gene copy number ($p=0.012$; $r=-0.324$, Pearson Test) and the promoter methylation status ($p=0.005$; $r=0.371$, Pearson Test) correlate significantly with ESR1 mRNA (ΔCt) transcript levels, suggesting that both mechanisms play an important role in the regulation of ESR1 gene expression in the NCI60 panel. No significant association between the copy number or methylation status with the ESR1 genotypes is noted (data not shown). To exclude these two potentially biasing regulatory factors, which are known to affect ESR1 gene expression, we are

interested in studying the differences in ESR1 mRNA expression for the genotypes at the ESR1 SNP locus in those cells that have both a normal gene copy number and an unmethylated promoter. However, because only 10 cell lines fulfill both requirements, a further stratification of those cells according to the genotypes would lack the statistical power to detect significant allelic differences between the ESR1 mRNA levels. Indeed, the data show an insignificant trend, whereby those 6 cells which harbor the T-allele of SNP rs2234693 express on average 2.86-fold higher mRNA levels than those 4 cells C/C in genotype (18.32 and 20.83 ΔCt, respectively; p=0.476, Mann-Whitney Test).

3.4.3. Discussion

In this Chapter, we have first described in detail the SNPs that are highly linked to the SNPs identified in our previous screens and found in the MDM4, PPP2R5E, YWHAQ, CD44 and ESR1 genes. To do this we utilized the data generated in the HapMap and 1000 Genomes projects. We are confident that for the regions we are focusing on the vast majority of the high-frequency variation has been identified in the projects. Specifically, for these genes, the 1000 Genomes project has successfully sequenced those regions in 179 individuals with an average mapped coverage of 3.6X per individual. Although an estimated discovery rate of SNP variants in the Pilot 1 of the project has not been given, the expected 95% coverage of polymorphic variation with an MAF of at least 1% with a slightly higher sequencing depth (above 4X) upon completion of the project in 1094 individuals, together with the extensive cataloguing efforts and a vast range of SNP discovery methods (e.g. gene-based studies, shotgun-sequencing,

targeted sequencing) already incorporated in the HapMap Project, suggests that the vast majority of high-frequency polymorphic variation has already been discovered (Durbin, Abecasis et al. 2010).

Once we had identified the linked SNPs we then explored if they resided in potential regulatory regions in order to identify candidate functional SNPs. None of the SNPs results in changes in the protein sequence. However, we have identified potential casual SNPs that could affect cellular protein levels through altering mechanisms such as transcription and microRNA regulation. We have then utilized a cell-line panel-based approach to search for allelic differences in gene expression patterns to begin to elucidate the molecular and functional regulatory mechanisms behind those SNPs.

3.4.3.1. Potential functional SNPs upstream of the second MDM4 promoter might regulate MDM4 gene expression in an allele-specific manner

We began our analysis with a haplotype in the MDM4 gene previously identified by our group that deviates from standard assumptions of selective neutrality and associates with a decreased risk for, and later onset of, breast and ovarian cancers (Atwal, Kirchhoff et al. 2009; Kulkarni, Vazquez et al. 2009). Our analysis shows that this selected haplotype comprises 14 closely linked SNPs in the MDM4 gene and demonstrates that it associates with lower MDM4 mRNA levels in the NCI60 cell line panel. Specifically, cells homozygous for the ancestral T-allele, which associates with an increased risk for, and earlier onset of ovarian and breast cancers (Atwal, Kirchhoff et al. 2009; Kulkarni, Vazquez et al. 2009), express significantly, on average 2.6-fold higher

MDM4 levels than those C/C in genotype, while the cells heterozygous at this locus are intermediate (p=0.031; Jonckheere-Terpstra Test; Figure 3.4.6). These data support a model, whereby the minor allele of one of the SNPs from the MDM4 haplotype leads through yet unidentified mechanisms to an increased MDM4 mRNA expression and protein levels. These elevated MDM4 expression could result in an increased activity of the MDM4 oncogene, which would lead to an attenuation of the ability of p53 to respond to cellular stress and, subsequently to reduced tumor suppression in cells. The analysis of putative regulatory regions suggests that polymorphisms from the selected MDM4 haplotype, particularly both SNPs located within approximately 5kb 5' of the second MDM4 promoter (Phillips, Teunisse et al. 2010) in intron 1 (rs1380576 and rs4951393), reside in a region of the MDM4 gene possibly involved in the regulation of transcription. This notion is supported by the fact that the SNPs associate with the expression of experimentally determined functional elements in the ENCODE project (Table 3.4.1), such as histone methylation patterns (H3K4me1, H3K4me3, H3K27ac), binding of sequence-specific transcription factors (Myc/Max) and components of the general transcription machinery (Pol II), which are indicative of active transcription and the presence of transcriptional start sites (Birney, Stamatoyannopoulos et al. 2007; Koch, Andrews et al. 2007; Rada-Iglesias, Bajpai et al. 2011).

Recently, a report has suggested that the microRNA hsa-miR-191 preferably binds to the minor ancestral C-allele of SNP rs4245739 in the 3'UTR of the MDM4 gene and regulates the steady-state MDM4 mRNA and protein expression in ovarian cancer cell lines and tumor tissues in an allele-specific manner (Wynendaele, Bohnke et al. 2010), whereby the direction of the allelic differences and significance of the findings is

inconclusive. Specifically, the authors reported a trend in a cell line panel comprising 9 cell lines, whereby cells homozygous for the minor C-allele express lower MDM4 protein levels than those with a T-allele. However, the authors note, surprisingly, that an opposite trend is observed on the mRNA level, whereby cells with a C/C genotype express relatively high MDM4 mRNA levels. Confusingly, one of the cell lines included in this cell line panel is the MCF7 cell line, which is, contrary to the claims of the authors, a breast and not an ovarian cancer cell line (Neve, Chin et al. 2006). Moreover, the authors fail to show a significant association of MDM4 protein levels and the rs4245739 genotype in tissues derived from ovarian carcinoma patients. In addition, in contrast to the finding in the cell line panel, the authors observe an association between higher MDM4 mRNA levels and the T/T genotype in ovarian cancer tissues. However, the authors do not mention which genotype(s) they compare these expression levels to and how they conduct the statistical analysis using a contingency table (Chi-square Test) instead of a test, which directly compares the means or medians of mRNA expression (Wynendaele, Bohnke et al. 2010).

Important to our study is the fact that the 3' UTR SNP rs4245739 analyzed by Wynendale et al. is not perfectly linked to the selected haplotype of the MDM4 gene (r^2=0.847; Table 3.4.1) and therefore the observed clinical and cellular associations of this SNP might differ from those of the other SNPs from the selected haplotype. To compare the effects of SNP rs4245739 on MDM4 mRNA expression in the NCI60 panel to the results observed with the SNP rs1563828 utilized in our study, which is more closely linked to SNP rs2290854 on which the selected haplotype is centered (r^2=0.960), we genotype SNP rs4245739 in the NCI60 panel. 45 out of the 59 cell lines have

identical genotypes and 15 differ in their genotype calls. Importantly, we do not observe any pronounced or significant association with SNP rs4245739 and MDM4 mRNA expression levels in all 59 cell lines or in the 6 ovarian cancer cell lines from the NCI60 panel (data not shown). This observation is in line with the inconclusive data presented by the authors regarding the significance and the direction of the regulation of MDM4 mRNA and protein levels in the mixed panel of cancer cell lines and ovarian tumor tissues (Wynendaele, Bohnke et al. 2010). Together, these data suggest that SNP rs4245739 and the other potential functional SNPs from the selected MDM4 haplotype might affect the activity of MDM4 through different and possibly independent mechanisms.

Taken together, our data suggest that the selected MDM4 haplotype affects the expression of steady-state mRNA levels of the MDM4 gene, which could lead to higher protein levels and oncogenic activity of MDM4, and an attenuation of the p53 stress response. This, subsequently, would result in a reduced ability of cells to undergo a malignant transformation and lead to the observed allelic differences in risk for, and later onset of, breast and ovarian cancers (Atwal, Kirchhoff et al. 2009). Furthermore, we identify potential causative SNPs in intron 1 of the MDM4 gene, located 5' of the second MDM4 promoter, which might reside in regulatory enhancer regions of the gene and thus affect the MDM4 expression in an allele-specific manner. Others in the group (Anna Grawenda, Alexander Finalyson) are currently working on elucidating the functional and molecular mechanisms in order to test this hypothesis.

3.4.3.2. The expression of PPP2R5E mRNA is possibly regulated by a selected haplotype in the PPP2R5E gene through a yet unidentified SNP

In Chapter 3.1, we have shown that PPP2R5E ε-SNP2 resides in an evolutionary selected haplotype in the PPP2R5E gene and associates with altered sarcoma risk and prognosis. Interestingly, we have found that cells homozygous for the minor allele of ε-SNP2, which associates with an earlier age of sarcoma diagnosis and increased overall risk for sarcoma in younger patients, express on average significantly, 1.47-fold higher PPP2R5E transcript levels compared to those with a C/C genotype (Figure 3.4.7). These data suggest that the haplotype linked to PPP2R5E ε-SNP2 affects the activity of PPP2R5E through an up-regulation of steady state mRNA levels. This in turn could lead to increased protein expression of this regulatory subunit of PP2A, which, upon speculation, could result in decreased dephosphorylation at regulatory amino acids of p53, such as serine 37, and an attenuated p53 response to DNA damage (Dohoney, Guillerm et al. 2004). The observation that both the age-dependent incidence as well as the overall sarcoma risk are particularly pronounced in male patients, suggests that gender-specific signaling might underlie those differences. Possibly, sex hormones might affect the transcription of the PPP2R5E gene and this regulation could be modified by ε-SNP2. Interestingly, we observe that the allelic differences in mRNA expression are significantly more pronounced in cell lines, which express high levels of androgen receptor, suggesting that androgen regulation of PPP2R5E expression might account for the observed allelic differences.

Taken together, these results support the notion that the selected haplotype in the PPP2R5E gene regulates the activity of the gene by affecting its gene expression that is dependent on the activity of sex hormone receptors. We identify 3 SNPs that are significantly linked to PPP2R5E ε-SNP2 in intron 2, which might potentially harbor enhancer elements (MacKenzie and Quinn 1999; Bharadwaj, Trainor et al. 2003; Klenova, Scott et al. 2004), and intron 3 of the gene. The analysis of conservation patterns and functional elements in the PPP2R5E gene does not identify a candidate functional SNP. However, although it is likely that the vast majority of high-frequency polymorphic variation has already been discovered in the HapMap and 1000 Genomes Projects (Durbin, Abecasis et al. 2010), it is possible that further SNPs linked to the haplotype will be emerge upon availability of the full 1000 Genomes dataset. In addition, the effects of a functional PPP2R5E SNP might be tissue and cell-type specific, and therefore it may not be possible to study those effects in the cell lines utilized in the ENCODE Project. Further sequencing efforts and a deeper characterization of functional elements are necessary to fine-map the genomic region surrounding the PPP2R5E gene and to identify regulatory elements within this region, in order to determine all variants linked to the selected haplotype and identify further potential candidate SNPs for molecular and functional analyses

3.4.3.3. Associations of chemosensitivity SNPs in the CD44 and YWHAQ genes with the gene expression in cells and identification of candidate functional SNPs

We were next interested in exploring whether allelic differences in gene expression might underlie the clinical and cellular phenotypes of the SNPs in the CD44 and YWHAQ genes identified in Chapter 3.2. Interestingly, we observe that the alleles of both SNPs, which associate with weaker responses to chemotherapeutic treatment, also show a trend for an association with higher mRNA transcript levels in their respective genes. This trend is significant for the CD44 SNP locus, whereby 4.35-fold higher mRNA transcript levels are observed in cells harboring the C-allele than in those homozygous for the T-allele ($p=0.021$, Mann-Whitney Test, Figure 3.4.8). These results indicate that the minor C-allele of CD44 SNP, or a minor allele of any other closely linked polymorphism, might lead through yet unidentified mechanisms to an elevated CD44 mRNA expression, possibly resulting in elevated CD44 protein levels and an increased oncogenic activity of this gene. Thus, upon treatment with DNA damaging chemotherapeutic agents, cells harboring the C-allele, which express higher CD44 levels, could more efficiently block stress-induced, p53-regulated apoptotic signals than those T/T in genotype and a lower CD44 expression (Godar, Ince et al. 2008).

We identify 7 SNPs in intron 1 of the gene, which constitute the haplotype centered on SNP rs187115. One of those SNPs resides in a highly conserved genomic region and 5 SNPs show some signs of regulatory potential in the ENCODE Project (Table 3.4.3), specifically the binding of Pol II to those genomic regions, indicating an association with potential TSSs. However, no transcriptional promoters have been reported in these intronic regions of the CD44 gene in the literature and no other signs of functional elements are reported in the ENCODE dataset, such as chromatin signatures indicative of transcriptional promoters or enhancers. Taken together, these data suggest

that one of those SNPs might possibly reside in an enhancer or intronic promoter region and affect the expression of the CD44 gene in an allele-specific manner, although more evidence for a regulatory role of the described genomic regions would be necessary to confirm their functional relevance.

In the case of the YWHAQ locus, however, the allelic differences in gene expression fail to reach statistical significance, both in all 59 cell lines ($p=0.088$, Jonckheere-Terpstra Test), as well as in the cell lines which retained a p53 wild-type gene, suggesting that other mechanisms, such as expression of other isoforms, alternative splicing or microRNA binding might be affected by this SNP and impact on the function of the YWHAQ gene. In addition, it is also possible that a SNP from the identified haplotype might regulate the expression or activity of another gene through distant cis- or trans-acting mechanisms. None of the 24 SNPs in introns 2, 3 and 5 as well as 3' of the gene, which constitute the extended haplotype, emerges as a clear candidate to affect any of the mechanisms that could potentially regulate YWHAQ. However, despite the relatively high estimated genomic coverage of the HapMap and 1000 Genomes Projects, it is possible that not all SNPs linked to the haplotype have been identified yet, specifically with the reported average sequencing depth of 3.6X in 179 individuals, or that the ENCODE Project does not provide enough information for the analysis of the regions in which the identified intronic SNPs reside. Further sequencing efforts and a deeper characterization of functional elements are necessary to fine-map the genomic region surrounding the YWHAQ gene and to identify regulatory elements within this region, in order to determine all variants linked to the selected haplotype and identify candidate functional SNPs. Furthermore, it is possible that, for example, SNP

rs114029751, which is moderately linked to SNP rs6734469 (r^2=0.741, Table 3.4.4), on which the studied haplotype is centred, is more closely linked to the true functional SNP. It would be interesting to genotype SNP rs114029751 in the cells from the NCI60 panel and determine whether any allelic differences in chemosensitivity profiles and YWHAQ gene expression exist for this locus.

3.4.3.4. Identification of potential functional SNPs that might affect ESR1 mRNA expression

In Chapter 3.3 we have described a linked SNPs in the ESR1 gene centered on SNP rs2234693, which associates with gender-dependent, allelic differences in the age of sarcoma and ovarian carcinoma diagnosis. Specifically, female, but not male, sarcoma as well as ovarian carcinoma patients, who carry the T-allele are diagnosed with tumors on average earlier in life than individuals C/C in genotype (Chapter 3.3). This observation is line with reports from the literature, whereby, similar to our findings, the T-allele associates with a higher risk for breast and endometrial cancer (Wedren, Lovmar et al. 2008; Anghel, Narita et al. 2010). Of the 17 linked SNPs, twelve are located in intron 1, four in intron 2 and one synonymous SNP in exon 1 of the ESR1 gene. The synonymous SNP rs2077647 is located 98bp from the nearest exon/intron boundary and therefore rather unlikely to affect splicing, as exonic splicing-regulating regions are thought to reside within 60bp of exon/intron boundaries (Blencowe 2000; Chew, Baginsky et al. 2000; Fairbrother, Yeh et al. 2002; Fairbrother, Yeo et al. 2004; Wang, Rolish et al. 2004; Szeszel-Fedorowicz, Talukdar et al. 2006; Wang and Burge 2008; Wang, Crystal et

al. 2009; Raponi, Kralovicova et al. 2011). However, this SNP is predicted to alter the binding sites of microRNAs hsa-miR-1296, hsa-miR-105*, and hsa-miR-222* (Figure 3.4.5). In similarity to a previous report, which showed that a synonymous SNP in the coding region of the IRGM gene can lead to allelic differences in microRNA binding and thus regulate gene expression (Brest, Lapaquette et al. 2011), the allelic differences in microRNA binding affinities at the ESR1 SNP^{rs2077647} locus have therefore the potential to result in allelic differences in ESR1 mRNA expression. Of the intronic SNPs, two polymorphisms reside in a well-conserved region flanking exon 2, which suggests a regulatory function for this region, and might constitute a distal enhancer in the ESR1 gene. To begin to explore potential regulatory mechanisms of these linked loci, we were first interested in investigating the possibility that these, or other SNPs might affect the activity of ESR1 by altering the expression of the gene. One recent study utilizing human brain tissue supports this notion, as allelic differences of this haplotype were noted in ESR1 mRNA levels (Weickert, Miranda-Angulo et al. 2008). Specifically, brain tissues from individuals harboring the T-allele of SNP rs2234693 had significantly increased ESR1 transcript levels compared to those with a C/C genotype (Weickert, Miranda-Angulo et al. 2008). However, the study of allelic differences in ESR1 expression is complicated by the fact that the expression of ERα is tightly regulated in human tumor cells, particularly by gene amplification or promoter methylation (Darbre and King 1988; Widschwendter, Siegmund et al. 2004; Holst, Stahl et al. 2007; Shen, Kondo et al. 2007). We have therefore determined that 44% of the NCI60 cell lines have a normal copy number and 41% of the cells have an unmethylated ESR1 promoter (Shen, Kondo et al. 2007) and only 10 cell lines (17%) have both, a normal copy number and an

unmethylated promoter. This level of complexity in the regulation of ERα expression in tumor cells makes it difficult to study or interpret the effects of the ESR1 SNPs on gene expression in the NCI60 human tumor cell line panel, as both regulatory factors are likely to bias the analysis. To reduce this potential bias, we have attempted to analyze the allelic differences in ESR1 gene expression in those 10 cell lines with both, a normal copy number and an unmethylated promoter. We only observe an insignificant trend in the NCI60 panel, whereby cell lines harboring the T-allele express on average 2.86-fold higher ESR1 mRNA levels than those C/C in genotype. It is, of course, possible that the observed differences in ESR1 levels in the NCI60 cell lines are insignificant because the SNPs do not regulate ESR1 transcription. However, the lack of a significant association might also be due to the low number of cells included in the calculations and the resulting lack of statistical power. In order to exclude the bias induced by the regulatory factors gene amplification and promoter methylation, we will utilize an isogenic system in Chapter 3.5 to further study the effects of the ESR1 SNPs on the gene expression.

In summary, the haplotype analysis and the study of steady state mRNA levels in the NCI60 cell line panel indicates that SNPs in the MDM4, PPP2R5E, CD44 and possibly ESR1 gene regulate the expression of the respective genes through yet unidentified molecular mechanisms, while the functional mechanisms of the YWHAQ SNP still remain elusive. Furthermore, the analysis of putative regulatory regions identifies candidate functional SNPs for further molecular and functional studies in the MDM4, CD44 and ESR1 genes. Importantly, the findings from this study need to be validated in other cellular systems and detailed functional studies need to be performed to test those hypotheses. Furthermore, this study is primarily based on the observation that

cells from the NCI60 panel associate with allelic differences in the steady state mRNA levels of the studied genes. Importantly, this model makes a major assumption: that the RNA levels mirror the protein levels in these cell lines. However, protein levels can be tightly regulated post-translationally under various experimental conditions (Toledo and Wahl 2007). For example, MDM4 protein levels are regulated through mechanisms such as the proteosomal degradation by the E3 ligase activity of MDM2 (Toledo and Wahl 2007). Therefore, it will be of considerable interest to study whether or not the measured mRNA levels of the investigated genes correlate with the protein levels in the NCI60 cell line panel, in order to directly assess whether or not cellular protein levels correlate with the genotypes of the identified SNPs.

The MDM4, PPP2R5E, CD44, YWHAQ and ESR1 genes are known to play important roles in regulating the p53 cellular stress response which, over the past thirty years, has been demonstrated to be crucial in suppressing tumorigenesis and mediating cellular responses to many cancer therapies (Sullivan, Syed et al. 2004). We hope that the results presented in this report will contribute to a better understanding of the molecular mechanisms associated with the genetic variants in the MDM4, PPP2R5E, CD44, YWHAQ and ESR1 genes, which have previously been shown to affect human cancer. Hopefully, this and subsequent work will help to gain further insight into the regulation of the expression of those genes in humans and, importantly, will eventually enable the optimization of cancer prevention and treatment strategies to patients.

3.5. Functional analyses of single nucleotide polymorphisms

3.5.1. Introduction

3.5.2. Results

3.5.2.1. ESR1 and MDM4 SNP loci associate with allelic imbalances in transcript levels in heterozygous cells

3.5.2.2. Functional analyses of a single nucleotide polymorphism in the estrogen receptor alpha gene

3.5.3. Discussion

3.5.3.1. An analysis of allele-specific regulation of ESR1 expression by SNP rs2234693

3.5.3.2. A selected haplotype in the MDM4 gene regulates MDM4 mRNA expression

3.5.1. Introduction

The haplotype analysis, the study of putative regulatory regions and the analysis of steady state mRNA levels in the NCI60 cell line panel, described in the previous Chapter, together indicate that SNPs in the MDM4, CD44 and possibly ESR1 gene regulate the expression of the respective genes. Furthermore, these analyzes highlight candidate functional SNPs for further molecular and functional studies. In this Chapter, we are interested in experimentally testing these predictions. First, we develop an assay, in which we can study the allele-specific transcript levels in heterozygous cells for a given SNP, in order to confirm our results from the cell panel analysis (Chapter 3.2.2.2) in an isogenic system. We utilize this system to measure allele-specific transcript levels of two of the genes, ESR1 and MDM4. Due to the results obtained, we focus on a candidate SNP in intron 1 of the ESR1 gene and explore its role in modulating the expression of estrogen receptor in cells.

3.5.2. Results

3.5.2.1. ESR1 and MDM4 SNP loci associate with allelic imbalances in transcript levels in heterozygous cells

In Chapter 3.3, we have identified a haplotype in the ESR1 gene that associates with allelic differences in age- and gender-dependent onset and risk of soft-tissue sarcoma development. Our observations, together with published reports from the literature, whereby the T-allele was shown to associate with a higher risk for breast and

endometrial cancer (Wedren, Lovmar et al. 2008; Anghel, Narita et al. 2010), leads to the hypothesis that the ESR1 haplotype harbors one or more functional SNP(s) that could alter the tumor promoting activity of the estrogen receptor in cells. As described in detail in Chapters 3.4.2.1.1 and 3.4.2.1.6, SNP rs2234693 is closely linked to 16 other SNPs in the HapMap and the 1000 Genomes Projects (r^2>0.8). Twelve of the SNPs are located in intron 1, four in intron 2 and one synonymous SNP resides in exon 1.

A previous study suggests that these linked SNPs might affect the activity of ESR1 by regulating its expression. Specifically, an association of the major alleles of the linked SNPs with higher ESR1 transcript levels was reported in a panel of human brain tissues (Weickert, Miranda-Angulo et al. 2008). To explore this possibility, we first measured steady-state ESR1 transcript levels in the NCI60 panel and analyzed the allelic differences in ESR1 mRNA expression in 10 cell lines, which did not show an additional regulation of ESR1 expression by gene amplification/deletion or promoter methylation (Chapter 3.4.2.2.4). Interestingly, we observed a trend in line with the study in human brain tissues (Weickert, Miranda-Angulo et al. 2008), whereby cell lines harboring the T-allele express on average 2.86-fold higher ESR1 mRNA levels than those C/C in genotype (Chapter 3.4.2.2.4).

We next chose to analyze and compare the allele-specific RNA levels of ESR1 in a cell line heterozygous for the linked SNPs, MCF7. In this isogenic system, we utilize a duplex real time RT-PCR reaction, whereby two differently fluorescence-labeled probes are specifically designed to target each allele of the exonic SNP (rs2077647; Tracker SNP) that is one of the linked SNPs (Figure 3.5.1A). The MCF7 cell line is an ER-α positive breast cancer cell line. We isolate genomic DNA and mRNA from

logarithmically growing cells. We generate a standard curve of the differences in fluorescence for both alleles over varying genomic DNA concentrations and utilize it to calculate the fold-difference in fluorescence for both alleles found in the cDNA synthesized from the RNA (Figure 3.5.1B). Interestingly, and in support of the previously reported observations, we observe, in this isogenic system, 2-fold more ESR1 RNA with the major alleles of the linked SNPs. Specifically, we note higher mRNA levels transcribed from the DNA strand containing the T-allele, as compared to the C-allele, of ESR1 SNP^{rs2234693} (Figure 3.5.1C).

The ability to detect and measure allelic imbalance of transcript levels in heterozygous cells has the advantage of excluding the biases inherent in cell line panel association studies, like we performed in Chapter 3.4.2.2. The cell panel we utilized is made up of cells that differ in many attributes other than the genotype of the studied SNP, therefore the differences in transcript levels noted for the cells of certain genotypes could have little to do with the allele. We, therefore, were interested to utilize this approach to further study the significant allelic differences in transcript levels we noted for an MDM4 haplotype in the NCI60 cell panel (Chapter 3.4.2.2.1). Importantly, the MDM4 haplotype is closely linked to an exonic SNP and therefore can be studied with our method.

We have shown in Chapter 3.4.2.2.1 that the selected haplotype in the MDM4 gene, which associates with allelic differences in breast and ovarian cancer risk (Atwal, Kirchhoff et al. 2009; Kulkarni, Vazquez et al. 2009), also associates with allelic differences in the expression of MDM4 mRNA in the NCI60 cell line panel. Hereby, cells harboring the minor alleles of the haplotype expressed significantly 2.6-fold higher

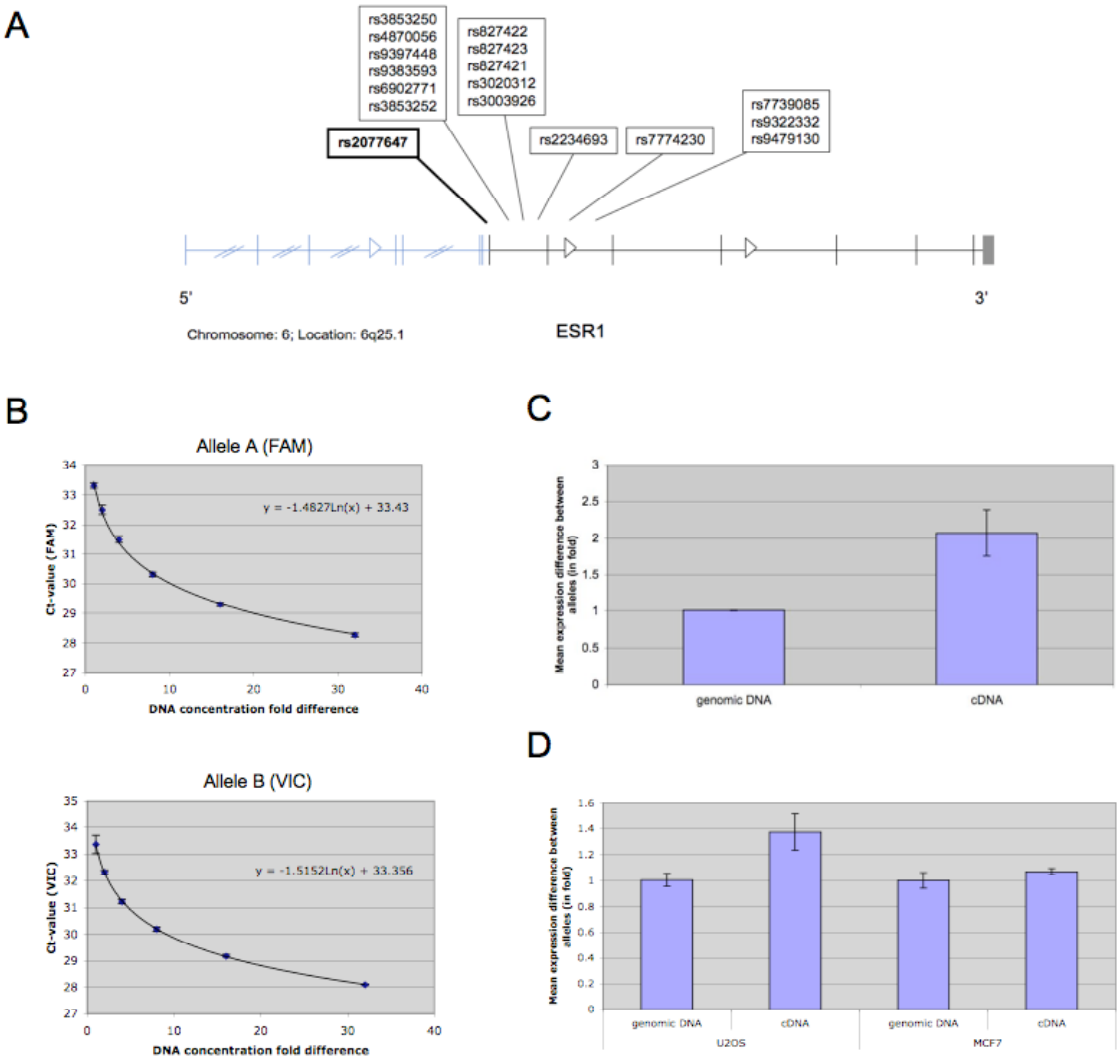


Figure 3.5.1. A quantitative, real-time PCR-based method to measure allele-specific transcription in an isogenic cellular system is developed and shows allelic differences in mRNA expression at ESR1 and MDM4 SNP loci. (A) List of SNPs in high linkage disequilibrium (r^2>0.8) with SNP rs2234693 in the ESR1 gene. The Tracker SNP in exon 1 is highlighted in bold. (B) A standard curve for the transcription of each of the two alleles is created based on the expression of a series of dilutions of the genomic DNA measurements of the same cell as the synthesized cDNA. Error bars represent the standard deviations of three independent readings. (C) Bar graph depicting the differences in the relative transcription of ESR1 mRNA from each allele of ESR1 SNP. Allele-specific quantification of ESR1 transcripts of non-treated MCF7 cells was performed and normalized to genomic DNA measurements from the same cells. A 2-fold increased transcription of ESR1 mRNA is measured from the strand containing the T-allele of SNP rs2234693 compared to the C-allele. (D) Bar graph depicting the differences in the relative transcription of MDM4 mRNA from each allele of MDM4 SNP normalized to genomic DNA measurements from the same cells. An up to 1.4-fold higher mRNA transcription is observed in the ERα-negative cell line U2OS from the minor T-allele of SNP rs1563828, while no differences are observed in the ERα-positive cell line MCF7. Error bars are depicted in all columns in (C and D) and represent standard deviations. 3 different measurements were performed (each 3 from 2 different mRNA extracts), each measurement was done in triplicate.

median transcript levels than those homozygous for the major alleles and the heterozygous cells expressed intermediate mRNA levels (Chapter 3.4.2.2.1; Figure 3.4.6). As the age-dependent cancer risk associated with this MDM4 haplotype has been shown to depend on the estrogen receptor status of the tumors, whereby the allelic differences were reported in patients whose tumors are ER-negative, but not ER-positive ((Kulkarni, Vazquez et al. 2009); Gareth Bond, unpublished results, personal communication), we have chosen one ER-negative (the osteosarcoma cell line U2OS) and one ER-positive cell line (MCF7) for the study of the allelic imbalance in transcript levels. Both cell lines are heterozygous for SNP rs1563828, which can distinguish the selected from the non-selected haplotypes, and for the exonic tracker SNP rs4245739 in the 3'UTR region of the gene (Chapter 3.4.2.1.1; Figure 3.4.1). We utilized a duplex real time RT-PCR reaction with the appropriate primers/probes as described above. In this isogenic system, we observed 1.4-fold more MDM4 RNA with the minor alleles of the haplotype in ER-negative U2OS cells, and no allelic differences in the ER-positive MCF7 cell line (Figure 3.5.1D). Specifically, in the U2OS cell line, we noted higher mRNA levels transcribed from the DNA strand containing the T-allele of SNP rs1563828, as compared to the C-allele (Figure 3.5.1D), in line with the observations in the NIC60 cell line panel. Together, these data support the hypothesis that one or more functional SNP(s) from the MDM4 haplotype regulate the expression of MDM4 transcripts in allele-specific and, possibly, ERα-dependent fashion.

3.5.2.2. Functional analyses of a single nucleotide polymorphism in the estrogen receptor alpha gene

We have chosen to focus our further analyses on the molecular and functional mechanisms, which regulate one or more of the linked ESR1 SNPs and lead to the noted changes in cancer risk and age-dependent incidence. The above-described data and analyzes suggest that the clinical associations observed for the ESR1 haplotype could be due to allelic differences in the transcriptional regulation of the gene. Estrogen receptor alpha has been shown to positively regulate the signaling of various pathways, such as the MAPK (Migliaccio, Di Domenico et al. 1996) and PI3K/AKT pathways (Simoncini, Hafezi-Moghadam et al. 2000), and to lead to the attenuation of the p53 stress response pathway (Okumura, Saji et al. 2002; Kinyamu and Archer 2003; Liu, Konduri et al. 2006; Sayeed, Konduri et al. 2007; Liu, Ip et al. 2009; Konduri, Medisetty et al. 2010). Indeed, heightened estrogen signaling can lead to a heightened proliferation in cells and an increased cancer risk (Migliaccio, Di Domenico et al. 1996; Simoncini, Hafezi-Moghadam et al. 2000; Kinyamu and Archer 2003; Liu, Ip et al. 2009). Therefore, alleles that increase expression of the estrogen receptor could result in heightened estrogen signaling and increase cancer risk (Figure 3.5.2). This model is consistent with our results thus far, that the major alleles of the linked SNPs associate with higher ESR1 mRNA and an earlier age of tumor onset.

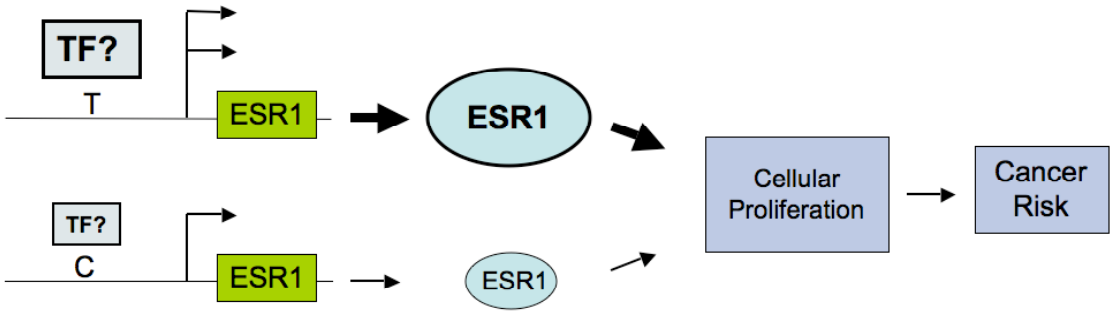


Figure 3.5.2. A model depicting a possible functional mechanism by which the identified SNPs in the ESR1 gene might affect estrogen signaling and cancer risk. A sequence specific transcription factor (TF) could bind with a different affinity to the sequences containing the alleles of one or more ESR1 SNP(s). This would result in an allele-specific activation of transcription of the ESR1 gene, leading to higher mRNA and protein levels associated with the major alleles. The resulting increased estrogen signaling could lead to an increased cellular proliferation and elevated cancer risk

As none of the 17 known linked SNPs reside in well-defined regulatory regions of the ESR1 gene, we first attempted to define putative regulatory regions by looking for evolutionary sequence conservation between species and functional regulatory elements in the ENCODE dataset (Chapter 3.4.2.1.6). Using two sequence alignment programs (CLUSTALW, MULAN) and three genomic databases (UCSC, VISTA, NCBI), we determined that only two of the 16 intronic SNPs, rs2234693 in intron 1 and rs7774230 in intron 2, reside in putative regulatory regions that are highly conserved from mouse to man on a scale of >100bp (Figure 3.5.3A). Of those two polymorphisms, only the region directly adjacent to and including SNP rs2234693 (+/- 6bp) shows a high level of conservation (>75%) (Figure 3.5.3B), while the region directly adjacent to SNP rs7774230 (+/- 6bp) shows no strong evidence of evolutionary conservation (<50% sequence similarity between human and mouse genomes). We, therefore, first chose to focus our analysis on SNP rs2234693, as the higher level of conservation of the region

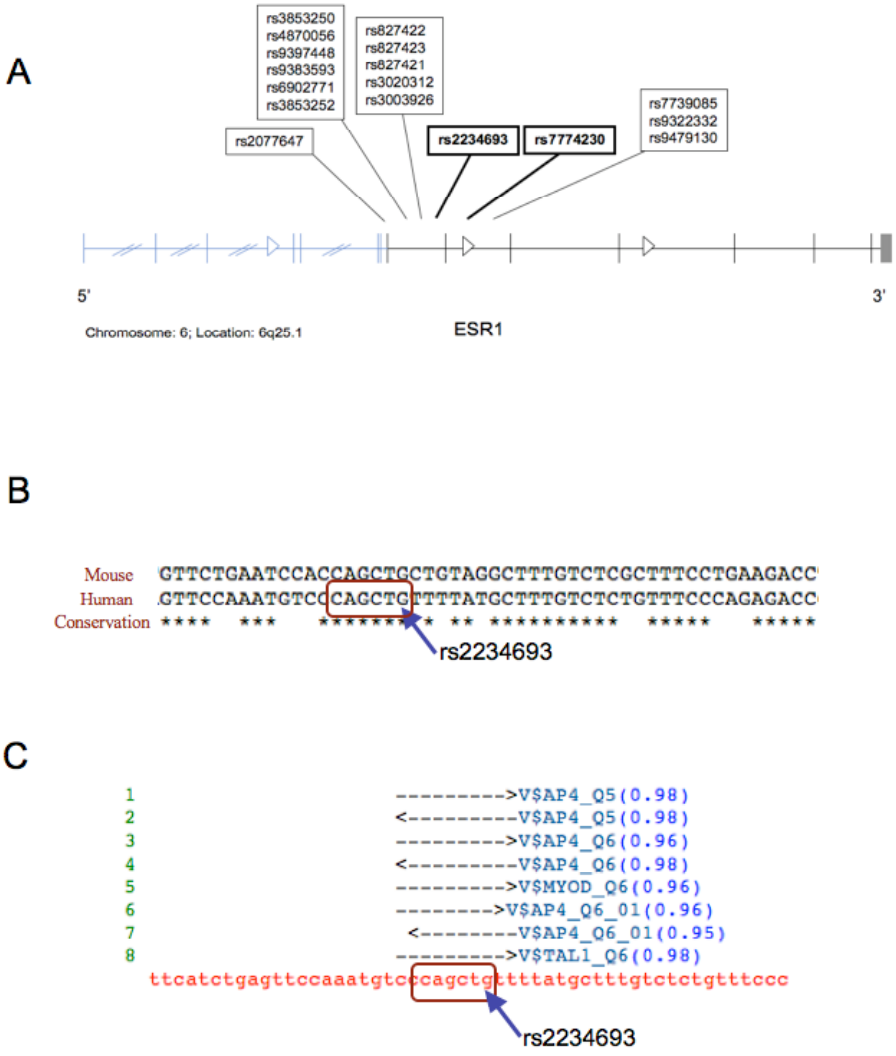


Figure 3.5.3. The ESR1 haplotype could harbor a SNP in a transcriptional enhancer element. (A) A list of SNPs in high linkage disequilibrium (r^2>0.8) with SNP rs2234693 of the ESR1 gene. Two of the SNP, which reside in conserved genomic regions are highlighted in bold boxes (B) An alignment of the nucleotide sequences of the human and mouse genomes comprising the sequence surrounding ESR1 SNP^{rs2234693}. The genomic region surrounding SNP rs2234693 (T-allele indicated by arrow) is evolutionary conserved from mouse to man, suggesting a regulatory function of this genomic region. The E-Box introduced by the T-allele is depicted in a red box. (C) A computational analysis predicts a high-affinity binding site of transcription factors from the basic Helix-Loop-Helix family to the sequence containing the T-allele.

immediately surrounding and including the SNP could indicate a putative regulatory region and a conserved transcription factor binding site. Indeed, the change from a C to the T residue introduces a consensus E-Box sequence (CANNTG; ESR1 SNP T-allele underlined; Figure 3.5.3B), which is typically bound by transcription factors from the basic-Helix-Loop-Helix (bHLH) family. Subsequently, a computational transcription factor binding site analysis (ExPlain 3.0; Transfac 2009.3) predicted the strongest allelic differences in the binding affinity for the bHLH transcription factors, specifically AP4, MyoD and Tal1, which were predicted to bind to the sequence harboring the T-allele, but not the C-allele of ESR1 SNP^{rs2234693} (Figure 3.5.3C).

In order to test these predictions, we performed a series of protein/DNA binding assays, electro-mobility shift assays (EMSAs), using identical oligonucleotides with either the T or C residues. First, we explored the ability of each oligonucleotide to compete the binding of the in vitro translated bHLH protein MyoD to a well-described E-Box DNA sequence (E2 E-Box; Figure 3.5.4A (Wang, Valdez et al. 2001; Buas, Kabak et al. 2010)). MyoD forms a heterodimeric complex with the bHLH protein E12, which enhances its ability to bind to the DNA (compare lane 2, 3, 4, Figure 3.5.4A; (Czernik, Peterson et al. 1996)). The binding of the MyoD/E12 complex to the E-Box is sequence-specific, as only an excess of unlabelled oligonucleotide carrying the wild-type E2 sequence (lane 5), but not the mutant sequence (lane 6) can compete the binding of the complex to the radioactively labelled oligo (Figure 3.5.4A). Consistent with the computational predictions, the oligonucleotides containing the T but not the C residue of SNP rs2234693 can abolish the MyoD/E12 binding to the E2 sequence (compare lane 7 to lane 8 of Figure 3.5.4A). Secondly, we explored if another bHLH protein showed

similar effects. Specifically, we assessed the binding of *in vitro* translated AP4 directly to each sequence. Interestingly, and consistent with the computational predictions and the MyoD/E12 competition experiment, AP4 bound with greater affinity to the T-allele containing DNA (compare lanes 2 and 6 of Figure 3.5.4B). The *in vitro* translated, flag-tagged AP4-DNA complex could be shifted with a flag-antibody, indicating that the observed bands were specifically caused by the AP4 protein binding to the DNA (lanes 3 and 7, Figure 3.5.4B). In addition, the complexes could be competed with an excess

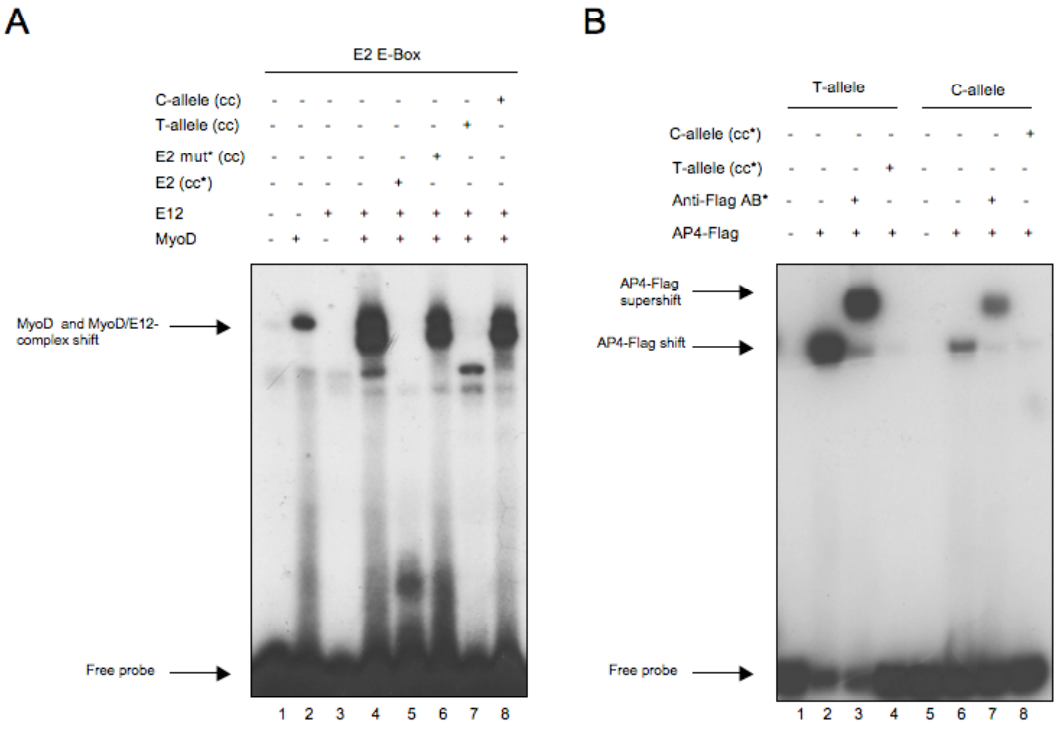


Figure 3.5.4. Electromobility shift assays (EMSAs) show a high binding affinity of bHLH transcription factors to the E-Box containing sequence with the T-allele of ESR1 SNP rs2234693. (A) The MyoD-E12 complex binds to a known E-Box containing sequence (E2 E-Box, lane 4), which is competed by the cold competitor T-allele (SNP T, lane 7), but not the C-allele containing sequence (lane 8). (B) A direct binding of exogenously overexpressed AP4 protein, tagged by a Flag-sequence, to the T-allele containing oligonucleotide (SNP T probe) is observed (lanes 2 and 3), while only a weak binding occurs to the C-allele-containing oligonucleotide (SNP C probe; lanes 6 and 7). *Abbreviations: cc, cold competitor; mut, mutant: AB, antibody

of unlabeled probes, indicating that the binding is sequence-specific (lanes 4 and 8, Figure 3.5.4B). Similar allelic differences in binding were seen with MyoD and Tal1 (data not shown).

We estimate that AP4 has up to 5-fold higher affinity for the T-allele than the C-allele. Specifically, using in-vitro translated AP4 protein, we performed EMSAs with increasing amounts of rabbit reticulocyte in the binding reaction (range 0.2-1.0µl of translated protein; Figure 3.5.5A). Hereby, we observed across the whole concentration range a stronger binding of the AP4 protein to the T-allele compared to the C-allele containing sequence, and, using a phosphoimager to quantify the relative band intensities, measured differences from 3 and up-to 5-fold difference between the binding to the two alleles (Figure 3.5.5A and B).

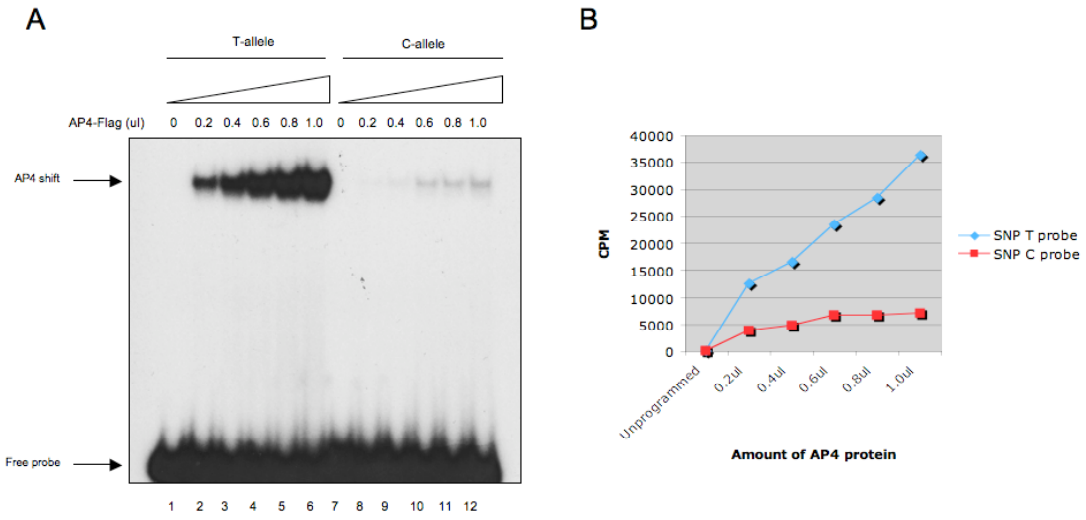


Figure 3.5.5. (A) An electromobility shift assay shows a high affinity of exogenously overexpressed AP4 protein to the E-Box containing sequence with the T-allele of ESR1 SNP^{rs2234693} (SNP T probe, lanes 1-6) across a wide protein amount range (0.2-1.0µl, while only a weak binding occurs to the C-allele-containing oligonucleotide (SNP C probe; lanes 7-12). (B) Quantification of the EMSA with a phosphoimager shows a stronger binding affinity of AP4 to the SNP T than to the SNP C containing sequence across the whole protein concentration range (up to 5-fold at 1.0µl of overexpressed protein). Abbreviations: CPM, counts per minute.

Next, we were interested in exploring if the noted allelic differences in bHLH binding could underlie the allelic differences we observed in transcript levels in the MCF7 cell line (Figure 3.5.1C). To begin to test this hypothesis, we performed the same above-described EMSAs, but used nuclear protein extracts from the MCF7 cell line, instead of *in vitro* translated proteins. Strikingly, and similar to our previous experiments, the T-containing DNA probe shifted to a significantly greater extent than the C-allele containing sequence (compare lanes 2 and 7, Figure 3.5.6A). The binding was sequence-specific, as indicated by the disappearance of the band with unlabeled wild-type (lanes 3, 8), but not mutant (lanes 4, 9) probe (Figure 3.5.6A). Importantly, 82% of the band

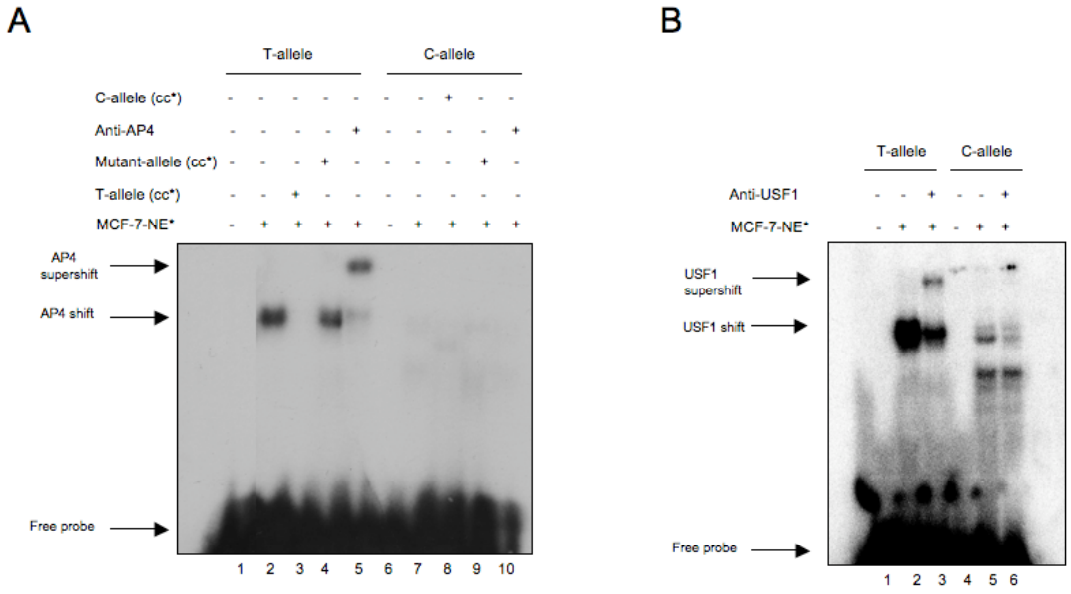


Figure 3.5.6. Endogenous AP4 and USF1 bind to the T-allele containing sequence of ESR1 SNP rs2234693. (A) An AP4-antibody supershifts endogenously expressed AP4 from MCF7 nuclear extracts bound to the SNP T allele containing sequence (lanes 2-6), while no binding is observed to the SNP C probe (lanes 7-10). (B) A part of the band representing a smaller complex of MCF7 nucelar extracts bound to the SNP T probe is supershifted by an USF1 antibody (lane 4). In contrast, this complex with the SNP C probe is hardly detectable (lanes 5-8). * Abbreviations: NE, Nuclear extracts; cc, cold competitor

shifted to a slower mobility upon incubation with an AP4-specific antibody, suggesting that AP4 binding to the T-allele containing probe is mainly responsible for the formation of the band (lanes 5, 10, Figure 3.5.6A). As another, less intense band could not be super-shifted with the AP4 antibody (lane 5, Figure 3.5.6A), we next explored the possibility that another bHLH transcription factor might bind to the oligonucleotide sequence in a T-allele-dependent manner. Indeed, the remaining protein-DNA complex (lanes 2 and 5, Figure 3.5.6B) was super-shifted with an antibody against the bHLH transcription factor USF1, specifically in the T-allele containing probe and less pronounced in the C-allele containing probe (compare lanes 3 and 6, Figure 3.5.6B).

In summary, we have shown *in vitro* that endogenous AP4 protein from the MCF7 cells can bind to the DNA sequence containing SNP rs2234693 in an allele-specific manner, thus possibly leading to the up-regulation of transcription specifically from the T-allele of this polymorphism. Next, we were interested in determining whether or not we could detect AP4 or USF1 on the genomic sequence surrounding SNP rs2234693 in the MCF7 cell line *in vivo* under normal cell culture conditions. We therefore performed a chromatin immunoprecipitation (ChIP) experiment in logarithmically growing cells. We first made sure that we could detect AP4 binding to a known enhancer region of the p21 gene using our antibody and conditions (Figure 3.5.7A) ((Jung, Menssen et al. 2008); Colin Goding, personal communication, unpublished data). The p21 promoter contains two adjacent E-Box elements and AP4 has been shown to utilize these sites to negatively regulate p21 expression ((Jung, Menssen et al. 2008); Colin Goding, personal communication, unpublished data). Importantly, we

clearly were able to amplify the two adjacent E-Box elements in the p21 promoter in the DNA isolated from the AP4-ChIP, but not from the IgG control (Figure 3.5.7A, compare lanes 1 and 2, as well as lanes 3 and 4 for the input DNA and non-template control, respectively). However, using the exact same conditions, we were unable to detect the presence of the sequence surrounding SNP rs2234693 in ESR1 (Figure 3.5.7B, compare lanes 1 and 2, as well as lanes 3 and 4 for the input DNA and non-template control, respectively).

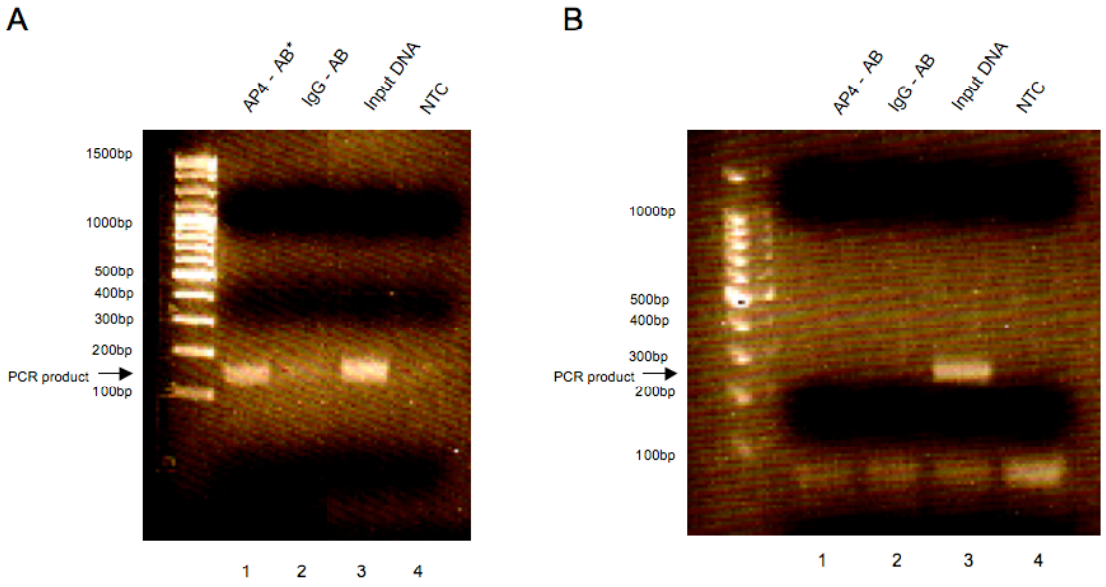


Figure 3.5.7. No significant binding of AP4 to the DNA sequence encompassing ESR1 SNP^{rs2234693} is observed *in vivo* in a chromatin immunoprecipitation (ChIP) assay in logarithmically growing MCF7 cells. (A) The binding of AP4 (lane 1) to the promoter region of p21, which is known to be regulated by AP4 via 2 E-Box elements, is demonstrated in this ChIP experiment, which served as a positive control. An unspecific IgG antibody (lane 2) raised in the same species as the primary antibody as well as DNA/RNA-free H_2O (lane 4) served as negative controls and the non-precipitated DNA (Input, lane 3) as a positive control for the experiment (B) In contrast, no significant binding of endogenous AP4 (lane 1) to the DNA sequence encompassing ESR1 SNP^{rs2234693} is observed. An unspecific IgG antibody raised in the same species as the primary AP4-antibody as well as DNA/RNA-free H_2O served as negative controls (lanes 2 and 4) and the non-precipitated DNA (Input, lane 3) as a positive control for the experiment The same MCF7 nuclear extracts were used as in (A). A 100kb-ladder DNA is depicted on the left side of the gels in (A) and (B).

These data do not support a role for AP4 in the allelic differences in estrogen receptor transcript levels. We therefore explored whether or not USF1 that we also found to bind to this region from the nuclear extracts in the EMSA (Figure 3.5.6B) can bind to the ESR1 SNP^{rs2234693} surrounding sequence *in vivo*. We performed a chromatin immunoprecipitation experiment in logarithmically growing cells, using the same experimental conditions as in the above-described ChIP assay. However, similar to AP4, we were not able to detect binding of USF1 to the regions surrounding ESR1 SNP^{rs2234693} (data not shown).

Together, these data do not lend support to the model, whereby the T-allele of rs2234693 leads to increased binding of the bHLH transcription factors AP4 and USF1, which results in an allele-specific up-regulation of transcription. However, we chose to further test this model utilizing an additional experimental approach. We decided to investigate the possible requirement for the bHLH proteins AP4 and USF1 in MCF7 cells for the allelic differences in estrogen receptor transcript levels. To this end, we first treated MCF7 cells with siRNA directed against AP4, as well as mock siRNA, and assessed the relative changes in mRNA expression from each allele using allele-specific RT-PCR. The AP4 siRNA was effective in reducing AP4 protein levels by 85% (Figure 3.5.8A and B). Furthermore, we were able to show that this reduction had a functional consequence, whereby a known AP4 target for suppression, p21, was up-regulated 4.5-fold (Figure 3.5.8C). Importantly, no significant effects were observed with the control siRNA (Figure 3.5.8A, B and C). Interestingly, reduction of AP4 levels significantly and repeatedly affected the allelic differences in ESR1 transcript levels. In particular, transfection of AP4 siRNA significantly decreased the transcript levels containing the T-

allele by 35% relative to the C-allele (Figure 3.5.8D). These data do support the predicted model, whereby AP4 preferably binds to the T-allele of SNP rs2234693 and up-regulates the transcription of ESR1 mRNA.

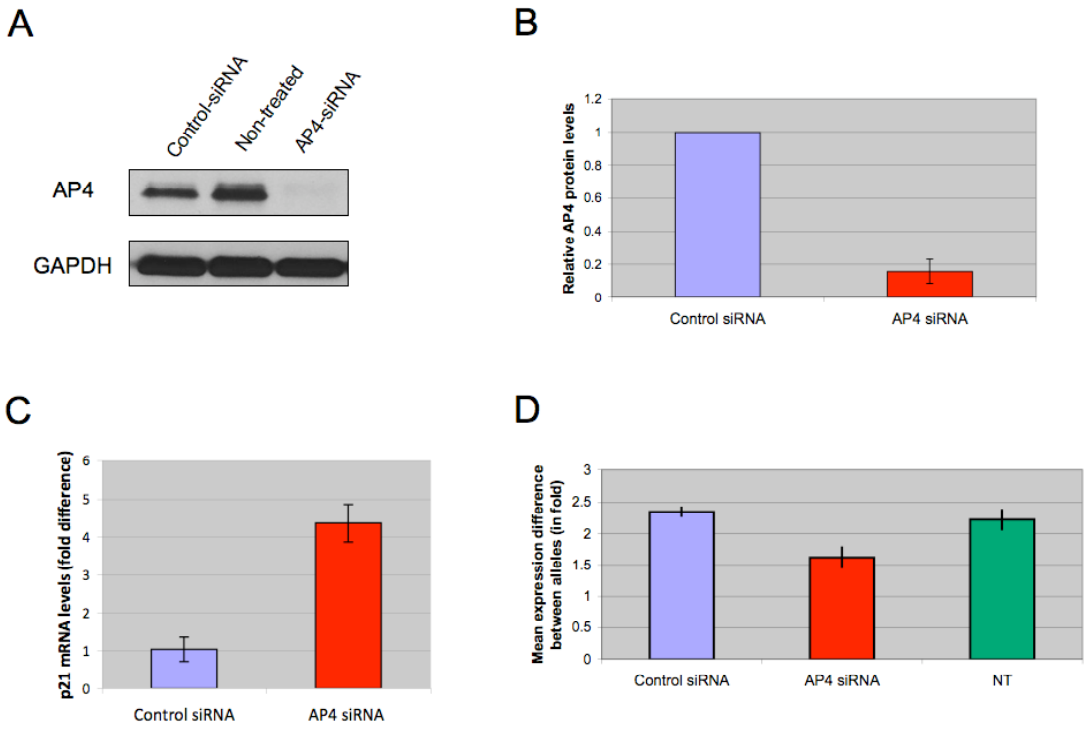


Figure 3.5.8. siRNA knock-down of AP4 in MCF7 cells results in a reduction of transcription from the T-allele of SNP^{rs2234693}. (A) Silencing of AP4 using a pool of 4 siRNAs directed against AP4 mRNA is effective in down-regulation of AP4 protein. (B) Quantification of the Western blot in (A). An 85% reduction AP4 protein expression is observed. (C) The effect of AP4 silencing on the expression of the mRNA of its known target gene p21, whose transcription increases more than 4-fold, is shown. (D) Allele-specific quantification of ESR1 transcripts of AP4- and mock-treated MCF7 cells showing a relative reduction of the transcription ratio of the T- versus the C-allele by 35% from over 2.3-fold with mock-siRNA-treated cells to approximately 1.6-fold with AP4-siRNA treated cells. All values are the averages of three independent readings of three independent experiments and the error bars represent the standard deviations.

We went on to investigate the possibility that USF1 also regulates the allele-specific expression of ESR1 mRNA by treating MCF7 cells with USF1 siRNA. The reduction of USF1 expression was successful on both the protein (Figure 3.5.9A) as well as the mRNA level (75% reduction of USF1 mRNA expression compared to mock siRNA, Figure 3.5.9B). However, no significant change in the relative expression from both ESR1 SNP alleles was noted (Figure 3.5.9C), suggesting that, in contrast to AP4, USF1 does not affect ESR1 mRNA transcription in an allele-specific fashion.

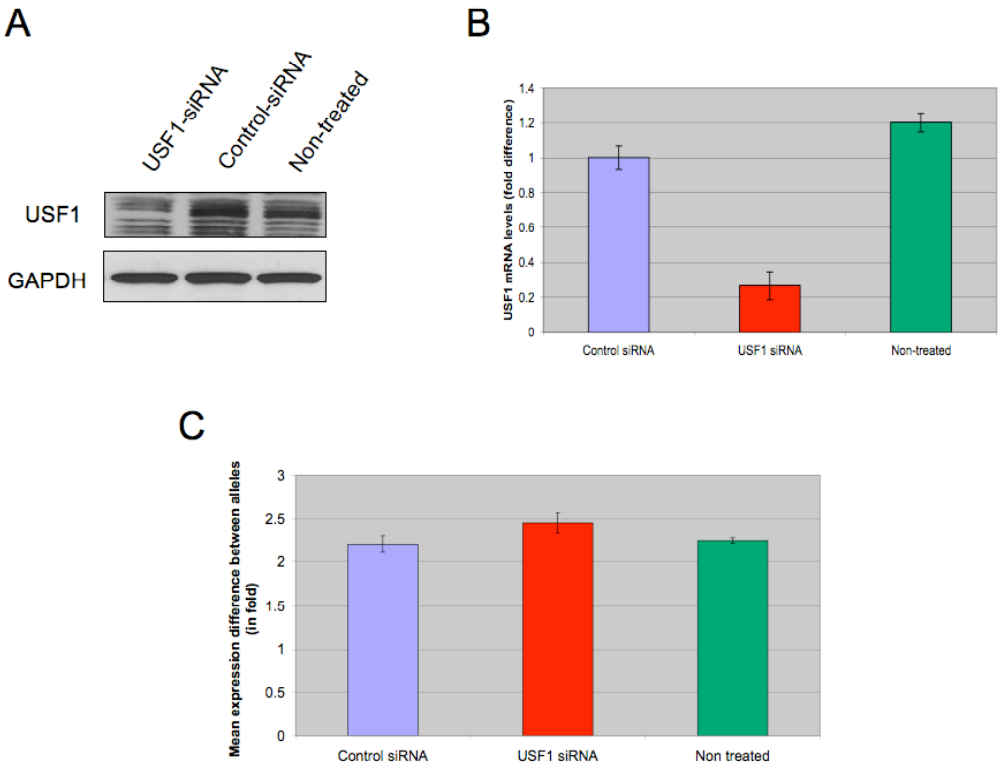


Figure 3.5.9. siRNA knock-down of USF1 in MCF7 cells does not affect the ratio of transcription between both alleles of SNP^{rs2234693}. (A) A Western blot showing an effective silencing of USF1 in siRNA knock-down experiments. USF1 corresponds to the dark bold band located between the unspecific dark bands (B) The specific reduction of USF1 expression is confirmed on the mRNA level by RT-PCR, whereby a 75% reduction is observed in USF1 siRNA treated cells (C) Allele-specific quantification of ESR1 transcripts of USF1- and mock-treated MCF7 cells shows no significant effect on the relative levels of allele-specific transcription of ESR1 mRNA. All values are the averages of three independent readings and the error bars represent the standard deviations.

AP4 has been suggested to enhance the binding of USF1 to the C/EBPα gene promoter and thus to affect adipocyte differentiation (Kim, Monila et al. 2007). Although the precise mechanism of this potential interaction is not known, we were interested whether both transcription factors can cooperate to regulate ESR1 mRNA expression in an allele-specific manner. To this end, we performed a simultaneous silencing of AP4 and USF1 expression using AP4 and USF1 siRNA and compared the effects of the double-knockdowns to the silencing with either one siRNA (co-transfected with the appropriate amount of mock-siRNA) or mock siRNA alone. Again, successful silencing of AP4 and USF1 was confirmed on the protein or mRNA level, respectively (Figure 3.5.10A and B). Similar to the previous observations, knockdown of AP4 alone resulted in a 40% reduction of the expression of the T-allele ESR1 mRNA compared to the C-allele ($p=0.0001$, T-test, two-sided), while no effect was observed with USF1 siRNA alone ($p=0.5142$, T-test, two-sided; Figure 3.5.10C). However, the simultaneous silencing of AP4 and USF1 slightly, but significantly, increased the relative expression of ESR1 mRNA originating from the T-allele strand compared to the AP4 single knock-down experiment ($p=0.0001$, T-test, two-sided), with a resulting 20% reduction of relative mRNA expression levels compared to mock siRNA (Figure 3.5.10C). These results suggest that USF1 can repress the relative expression of ESR1 mRNA from the T-allele containing DNA strand upon reduction of AP4 protein levels in cells.

To further test the dependence of AP4 on the allele-specific expression of estrogen receptor, we wanted to perform similar experiments in multiple cell lines with all three different genotypes of SNP rs2234693. To do this, we identified a panel of

breast cancer cell lines reported to be ERα positive (Neve, Chin et al. 2006), which included BT-474 (C/C in genotype for SNP rs2234693), CAMA1 (C/C), MCF7 (T/C), T47D (T/T) and ZR75/1 (T/T). We measured the steady-state mRNA levels of ESR1 and

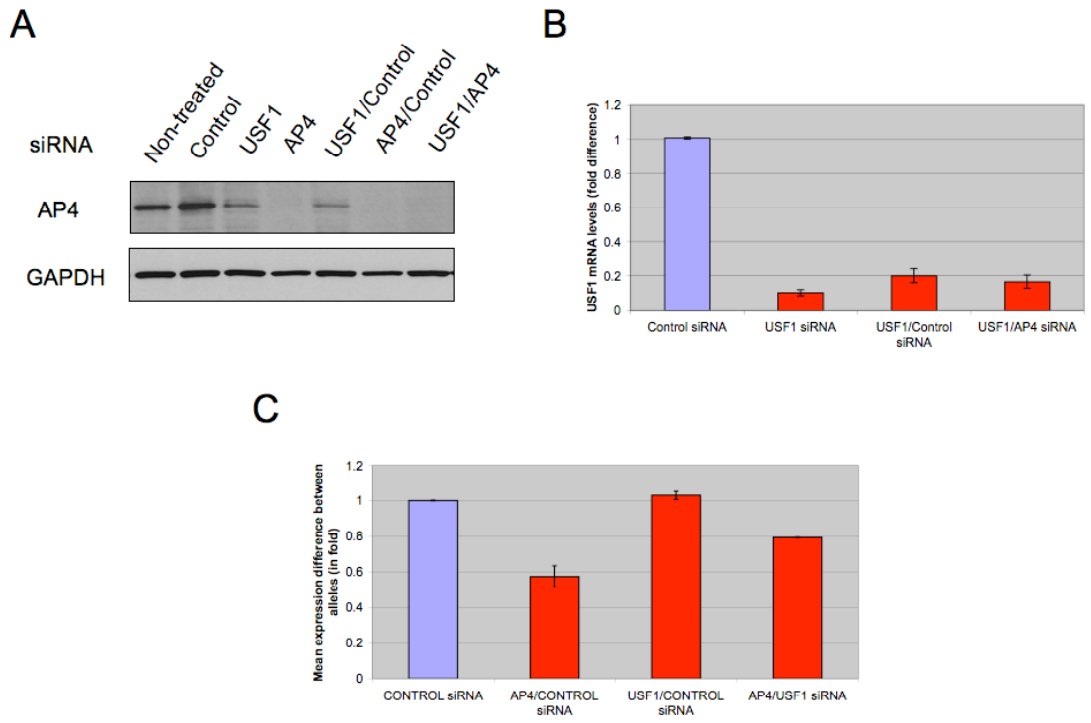


Figure 3.5.10. Simultaneous siRNA knock-down of AP4 and USF1 in MCF7 cells does affect transcription from the T-allele of SNP^{rs2234693}. (A) A Western blot showing an effective knock-down of AP4 in single and simultaneous siRNA knock-down experiments of the gene. (B) The specific reduction of USF1 expression is confirmed on the mRNA level by RT-PCR measurements, whereby an up-to 90% reduction is observed in USF1 siRNA treated cells (C) An approximately 40% reduction of the transcription from the T- versus the C-allele is observed upon treatment with a combination of AP4- and mock-siRNA, while no effect is observed upon combined treatment with USF1- and mock-siRNA. Upon combined treatment with AP4 and USF1 siRNA an intermediate phenotype is observed with a modest reduction of the T- to C-ratio. All values are the averages of three independent readings and the error bars represent the standard deviations.

protein levels of ERα, AP4 and USF1 in all cell lines. Interestingly, the two cell lines homozygous for the C-allele that in the MCF7 cell line associated with a significantly lower ESR1 mRNA expression compared to the T-allele, showed the lowest ESR1 transcript levels, whereby the ESR1 transcript was not even detectable in CAMA1 (C/C) (Figure 3.5.11A). In contrast, the cell lines either homozygous or heterozygous for the T-allele expressed 12-fold (MCF7; T/C), 14-fold (T47D; T/T) and 4-fold (ZR75-1; T/T) more steady-state ESR1 mRNA than the BT-474 (C/C) cell line (Figure 3.5.11A).

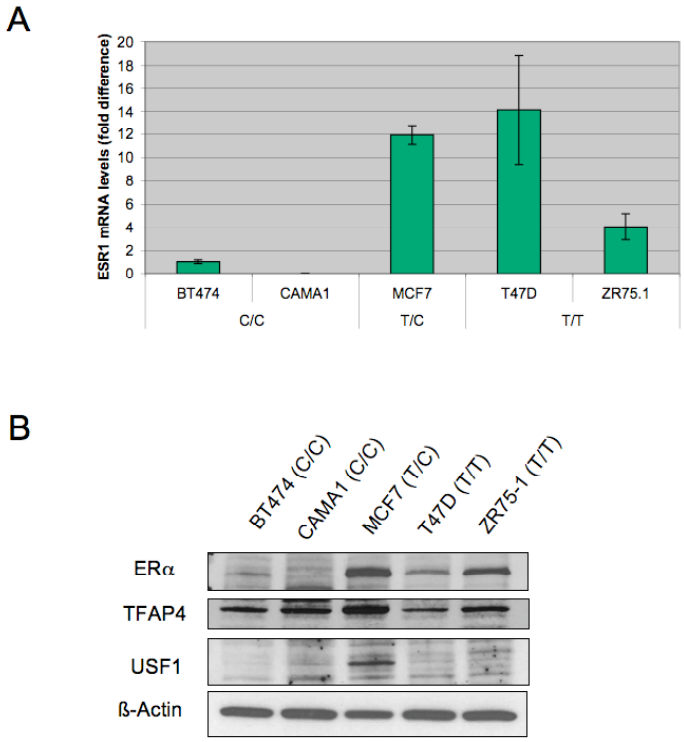


Figure 3.5.11. High gene expression levels of ESR1 associate with the T-allele of SNP rs2234693 in a panel of 5 breast cancer cell lines. (A) Bar graph showing the fold-difference in expression of ESR1 mRNA as measured by real-time RT-PCR (normalized to GAPDH levels). Cell lines harboring the T-allele of SNP rs2234693 express high ESR1 mRNA levels compared to those C/C in genotype. All values are the averages of three independent readings of two independent cell lysates and the error bars represent the standard deviation for each cell line value. (B) Western blot analysis of ERα protein levels and its potential upstream regulators AP4 and USF1. The lower panel is a Western blot of the same filter using a monoclonal antibody to ß-actin.

To extend these observations to ERα protein, we examined whole cell protein extracts from each cell using western blotting and similar trends were observed. As seen in Figure 3.5.11 panel B, no significant ERα was detected in the CAMA1 cell line (C/C). The second-lowest ERα levels were observed in BT-474 (C/C), and the highest protein levels were noted in the three remaining cell-lines that harbor at least one copy of the T-allele (Figure 3.5.11B).

Together, these data support our observations made in the heterozygous cell line, MCF7, that the T-allele results in higher ESR1 transcript levels, and extends them to estrogen receptor protein levels. We next wanted to test the dependence of AP4 on the association of the T-allele and higher estrogen receptor RNA and protein levels. We first determined that all cell lines expressed AP4 protein, with the highest protein levels observed in the MCF7 cell line (Figure 3.5.11B). Subsequently, we silenced the expression of AP4 with siRNA in all 4 ESR1 expressing cell lines (BT-474 C/C; MCF7 T/C; T47D T/T; ZR75-1 T/T) using the same experimental conditions as our previous experiments with MCF7 alone (Figure 3.5.12). The AP4 levels were successfully reduced in all 4 cell lines (Figure 3.5.12A). We aimed to confirm the efficient silencing of AP4 by measuring p21 mRNA levels. We observed a 2- to 6-fold increase in p21 transcript levels in the T47D, ZR75-1 and MCF7 cell line (Figure 3.5.12B), in line with the role of AP4 as a transcriptional silencer of p21 expression (Jung, Menssen et al. 2008). However, in the BT-474 cell line, we observed the opposite effect of AP4 siRNA on p21 expression, whereby p21 mRNA levels were reduced by over 50% (Figure 3.5.12B).

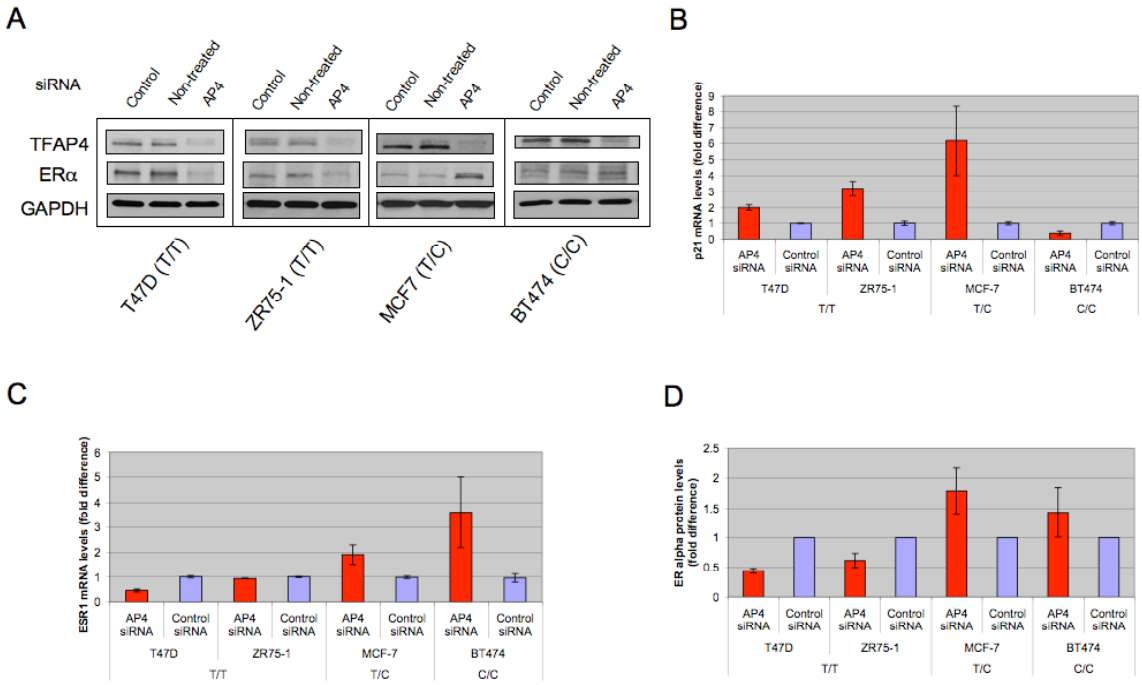


Figure 3.5.12. Silencing of AP4 with siRNA down-regulates ESR1 protein and mRNA expression in cells homozygous for the T-allele of SNP rs2234693 and up-regulates the expression in those harboring the C-allele. (A) A Western blot analysis of ERα protein levels and its upstream regulator AP4 upon treatment of cells with AP4 or mock siRNA. The lower panel is a Western blot of the same filter using a monoclonal antibody to GAPDH. (B) In all cell lines, a significant effect on the mRNA expression of the known AP4-target gene p21 was observed. (C) A bar graph showing the expression of ESR1 mRNA as measured by real-time RT-PCR (fold change relative to control siRNA treated cells; normalized to GAPDH). One of the two cell lines homozygous for the T-allele (T47D) shows an over 50% reduction of overall ESR1 mRNA levels, while the heterozygous cell line MCF7 associates with a 2-fold increase of ESR1 mRNA expression. The BT474 C/C cells show the strongest, 3-fold increase of ESR1 mRNA levels. (D) A quantification of the Western blot in (A) showing the effect on ERα protein levels after silencing of AP4 expression. The two cell lines homozygous for the T-allele show a reduction of overall ERα protein levels, while the heterozygous cell line MCF7 associates with a 2-fold increase of ERα protein expression. The same, although non-significant trend is observed in the BT474 cell line, which expresses only low levels of ERα protein. All values are the averages of three independent readings and 2 independent experiments and the error bars represent the standard deviations.

Upon AP4 siRNA treatment, we observed significant differences in the changes of ESR1 mRNA expression between the cells with different genotypes at the ESR1 $SNP^{rs2234693}$ locus. Specifically, AP4 reduction had the following effects on ESR1 transcript levels. The strongest (over 50%) reduction of ESR1 mRNA was observed in the T47D cell line (T/T in genotype), followed by unaltered mRNA expression in the ZR75-1 cell line (T/T in genotype), a 2-fold induction of ESR1 mRNA expression in the heterozygous MCF7 cell line and a relatively strong 3.5-fold ESR1 transcript induction in the BT-474 cell line, which is homozygous for the C-allele ($p=0.0001$, ANOVA-Test; Figure 3.5.12C). These results are potentially in line with the observation made in the heterozygous cell line, MCF7, that AP4 silencing reduces the relative expression of the T-allele of ESR1 $SNP^{rs2234693}$.

Next, we assessed whether the changes in ESR1 mRNA levels upon silencing with AP4 siRNA result in changes in ERα protein expression. Hereby, both cells homozygous for the T-allele showed a significant reduction of ERα protein levels, whereas in the heterozygous MCF7 cells, the overall effect on ERα expression was an increase in protein levels, similar to the BT-474 cell line (C/C in genotype), where a slight trend for an increase of ERα protein levels was observed (Figure 3.5.12A and D).

We next aimed to study the role of USF1 in the regulation of ESR1 expression in the breast cancer cell line panel, as the simultaneous silencing of USF1 and AP4 in the heterozygous MCF7 cell line affected allele-specific ESR1 mRNA expression (Figure 3.5.10). However, in the breast cancer cell line panel, only the MCF7 cell line expressed significant levels of the USF1 protein (Figure 3.5.12B) and we were therefore unable to further study the role of USF1 in this panel. Together, we have observed in the cell panel

that silencing of AP4, which is expressed in all 4 studied cell lines, affects the expression of ESR1 mRNA and protein in an allele-specific manner, whereas the role of USF1 can not be further studied in the panel, as it is only expressed in the MCF7 cell line.

In order to further analyze the regulatory mechanisms by which AP4 affects the expression of ESR1 mRNA in an allele-specific manner, we aimed to explore whether the conserved region surrounding SNP rs2234693 constitutes a transcriptional enhancer region. To do this, we PCR-amplified the evolutionary conserved region surrounding SNP rs2234693 (465 base-pairs in length) and cloned it into a pGL3 luciferase vector under the control of the thymidine kinase promoter (pGL3-TK). We did this for each allele of the SNP and confirmed the sequence identity of the two plasmids, except for the polymorphic base pair, through direct sequencing. We utilized these vector plasmids in transfection experiments into HeLa cells. The vectors were titrated in up-to a total of 500ng. However, despite a moderate, approximately 2-fold enhancement of the luciferase activity with the inserted sequence as compared to the empty promoter vector, no allele-specific enhancement was observed (Figure 3.5.14A), failing to provide support to the idea that the presence of the DNA strand containing the T-allele of ESR1 SNP^{rs2234693} constitutes a stronger enhancer than the C-allele containing sequence. Furthermore, no allele-specific transactivation was observed in the same experimental approach in COS and MCF7 cell-lines (data not shown).

Subsequently, we co-transfected the reporter plasmids with an AP4 expressing pCMV vector to help assess whether the enhancer function depends on the activity of the AP4 protein. A pGL3-p21-promoter vector served as a positive control, which carried two consensus E-Box sequences. Upon co-transfection of the pGL3-p21-promoter vector

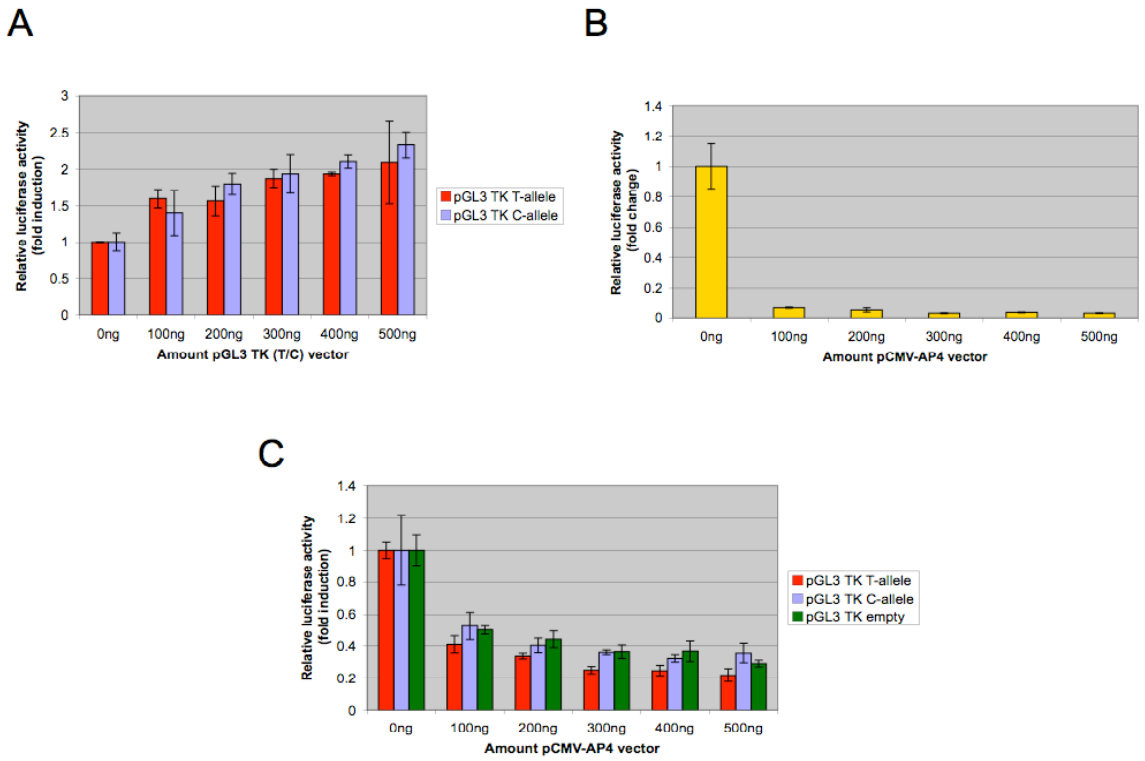


Figure 3.5.13. Luciferase reporter assays do not show an AP4-dependent allele-specific enhancer function of the conserved region surrounding ESR1 SNP[rs2234693]. (A) Despite a moderate, approximately up to 2-fold enhancement of the luciferase activity with the inserted sequences and increasing plasmid amounts as compared to the empty pGL3-TK promoter vector, no allele-specific enhancement was observed. (B) The over 10-fold repression of luciferase activity of the pGL3-p21 promoter vector (300ng of plasmid) by the AP4-expressing pCMV vector (as compared to the empty pCMV vector) served as a positive control for the experiment. (C) No significant allele-specific induction of luciferase activity as compared to the empty pGL3-TK vector plasmid was observed when the plasmids carrying the allele-specific sequences were co-transfected with an AP4 expressing pCMV vector. The total amount of the inserted vector was kept constant at 500ng in (A) and at 800ng in (B) and (C).

with increasing amounts of the AP4-pCMV vector (up to 500ng total, together with a corresponding amount of an empty pCMV vector , to keep the total amount of transfected pCMV plasmid constant at 500ng) a strong repression of the luciferase activity was

observed (Figure 3.5.14B), in line with the role of AP4 as a suppressor of p21 transcription (Colin Goding, unpublished results, personal communication). However, when co-transfected with a pGL3-TK vector carrying the ESR1 sequence and the different SNP-alleles, AP4 expression did not result in a significant change in luciferase activity compared to the empty pGL3-TK vector (Figure 3.5.14C). These data fail to provide support to the hypothesis that this region constitutes an AP4-dependent enhancer. Similar results were observed in the COS and MCF7 cell lines (data not shown). Furthermore, similar results were obtained with a pGL3-TK vector containing a triple repeat sequence with either the T- or C-allele, aimed to increase the sensitivity of the luciferase system to detect allelic imbalance (data not shown).

3.5.3. Discussion

3.5.3.1. An analysis of allele-specific regulation of ESR1 expression by SNP rs2234693

In Chapter 3.3, we identified two linked SNPs (rs2234693 and rs9479130) in the ESR1 gene, which associate with an age- and gender-dependent risk for early cancer onset in three independent cohorts. Next, in Chapter 3.4 (Sections 3.4.2.1.1, 3.4.2.1.6, 3.4.3.4), we have identified 15 additional SNPs, which are closely linked to SNPs rs2234693 and rs9479130. Of the 17 SNPs, twelve are located in intron 1, four in intron 2 and one synonymous SNP in exon 1 of the ESR1 gene. Of the intronic SNPs, two polymorphisms (rs2234693 and rs7774230) reside in well-conserved intronic regions, which might constitute a distal transcriptional enhancer in the ESR1 gene.

In this Chapter, we first chose to study whether one or more major alleles of the identified ESR1 SNPs might up-regulate the gene expression of the ESR1 gene, as suggested by a recent study with human brain tissues (Weickert, Miranda-Angulo et al. 2008) and our observations in the NCI60 cell line panel (Chapter 3.4.2.2.4). Interestingly, and in line with all previous observations, we observed both in the heterozygous MCF7 cell line and in a breast cancer cell panel a significantly, higher steady state ESR1 transcript levels associated with the T-allele of ESR1 SNP^{rs2234693} (Figures 3.5.1C and 3.5.11) Together, these results supports the hypothesis that the haplotype alleles linked to the T-allele of ESR1 SNP^{rs2234693} result in relatively high ESR1 mRNA and protein expression in cells. Based upon these preliminary results, it will be important to employ a larger panel of ER-positive cells in order to validate the observations. Furthermore, it will be necessary to determine the promoter methylation and gene amplification status of the ESR1 gene in these cells, in order to exclude the possibility that the observed associations, such as the lack of ESR1 expression in the CAMA1 (C/C) cell line, are affected by SNP-independent regulatory mechanisms, as cell lines have been shown to change the methylation patterns of promoters, including the estrogen receptor, during culture propagation (Antequera, Boyes et al. 1990; Ottaviano, Issa et al. 1994; Lai, Cheng et al. 2005).

We subsequently focused our study on the ESR1 SNP^{rs2234693} that resides in a highly conserved genomic region, and could therefore potentially alter the function of an intronic enhancer, which would affect ESR1 mRNA transcription (Figure 3.5.3B). Our results support this hypothesis in part. Specifically, our *in silica* and *in vitro* analyses show that the T-allele of ESR1 SNP^{rs2234693} introduces an E-Box sequence, which can be

bound by the basic Helix-loop-helix transcription factors AP4 and USF1 (Figures 3.5.3C, 3.5.4-3.5.6). Both bHLH transcription factors are known to affect the expression of genes involved in mitogenic and/or apoptotic signaling in cells, e.g. through transcriptional regulation of c-Myc, c-Fos, Caspase 9 and FOXE1 (Runkel, Shaw et al. 1991; Tsujimoto, Ono et al. 2005; Jung, Menssen et al. 2008; Landa, Ruiz-Llorente et al. 2009). In line with the *in silica* and *in vitro* results is also the observation that silencing of AP4 expression reduces the expression of the T-allele relative to the C-allele by 35-40% in the heterozygous MCF7 cell line, suggesting that AP4 can regulate ESR1 transcription in an allele-specific manner (Figure 3.5.8).

The observation that simultaneous silencing of AP4 and USF1 leads to an increased expression from the T-allele compared to the silencing of AP4 only, while no allele-specific effects are observed upon single silencing of USF1, suggests that USF1 might act as a weak repressor of transcription from the DNA strand linked to the T-allele of ESR1 $SNP^{rs2234693}$ (Figure 3.5.9 and 3.5.11). This could be achieved either through a direct competition with AP4 for the DNA-binding site(s), or through a direct interaction between AP4 and USF1. However, the precise mechanisms of such a potential interaction between AP4 and USF1 are not known, as although AP4 has been suggested to enhance the binding of USF1 to the C/EBPα gene promoter, it has also been shown to affect transcription only by forming homodimers (Hu, Luscher et al. 1990; Kim, Monila et al. 2007)

Despite the fact that we observed allelic differences in ESR1 mRNA and protein expression after AP4-silencing in the breast cancer cell line panel, these results are difficult to interpret. Specifically, in the BT-474 cell line, we observed a 50% decrease of

p21 expression upon treatment with AP4 siRNA (Figure 3.5.12B), which is in conflict with the role of AP4 as a transcriptional repressor of p21, as previously shown in the MCF7 cell line (Jung, Menssen et al. 2008). This might possibly be due to cell type-specific regulation of p21 transcription by AP4, but could also indicate that AP4 itself is inherently different in these cells. Therefore, a direct comparison of the effect of AP4 silencing on ESR1 expression cannot be made in these cell lines at this stage. Based on these preliminary results, additional controls for the effects of AP4 silencing should be utilized, and, subsequently, these effects examined in a larger ER-positive cell line panel and in additional isogenic systems, before a conclusion can be drawn from the data that AP4 possibly regulates ESR1 expression in an allele-specific manner. In addition, the fact that silencing of AP4 increases overall ESR1 mRNA and protein levels in the MCF7 cell line does not support the model, whereby AP4 acts as a transcriptional activator of ESR1 expression and thus leads to increased estrogen signaling in cells.

Moreover, the fact that we have not been able to show the binding of endogenous AP4 to the genomic region surrounding $SNP^{rs2234693}$ in the ChIP assay, clearly does not support a role for this SNP in the regulation of gene expression by AP4 (Figure 3.5.7). However, there are many possible explanations, why the ChIP technique could fail to show the binding of the transcription factors to the genomic region despite it being a true enhancer region. For example, in contrast to the p21 promoter, the region surrounding $SNP^{rs2234693}$ might constitute only a weak enhancer, which might be left undetected by the ChIP technique. This would be particularly apparent in a heterozygous cell line with only the T-, but not the C-allele being readily amenable to binding by the bHLH transcription factors. It would be interesting to perform additional ChIP assays in cells homozygous for

the T-allele of SNP^{rs2234693}, such as T47D or ZR75-1, in order to test this possibility. In addition, the epitope of the AP-4 protein, which is detected by the antibody used in our experimental set-up, might be masked by other transcriptional co-factors involved in the regulation of ESR1 expression, like USF1. The utilization of another antibody that recognizes a different epitope would be useful in ruling out this possibility.

Furthermore, we have not been able to detect any allele-specific enhancement of luciferase activity in reporter assays, both with and without co-transfection of an AP4-expressing plasmid (Figure 3.5.14). Although it seems unlikely, the reporter assays might also lack the sensitivity to detect subtle effects of AP4 regulation in this genomic region. In addition, the function of the potential enhancer region surrounding SNP rs2234693 might be dependent on other regulatory elements or DNA sequences unaccounted for in our experimental set up using artificial reporter constructs. However, it would be interesting to use an USF1-expressing vector in the co-transfection experiments to test whether USF1 might possibly affect the function of this potential enhancer region as suggested by our analysis.

Taken together, the functional and molecular analysis of the ESR1 haplotype suggests that one or more functional SNP(s) from this haplotype regulate ESR1 mRNA expression in an allele-specific manner. Hereby, through a yet unidentified mechanism, one or more major alleles linked to the T-allele of SNP^{rs2234693} could lead to an up-regulation of ESR1 mRNA expression and protein levels in cells and to heightened proliferation and increased cancer risk. Our data indicate that the bHLH transcription factor AP4 might be involved in such allele-specific regulation of ESR1 expression. However, in sum, our data does not indicate that AP4 regulates ESR1 transcription

through direct binding to the region surrounding ESR1 SNP^{rs2234693}. Possibly, another yet unidentified polymorphism closely linked to ESR1 SNP^{rs2234693} affects the DNA binding affinity of AP4, or this bHLH factor regulates ESR1 expression through other, indirect mechanisms. It will be of great interest to precisely determine experimentally whether the genomic regions, in which the 16 other linked ESR1 SNPs reside, harbor elements involved in the transcriptional regulation of the ESR1 gene. To this end, particularly the determination of chromatin accessibility, histone modification patterns and binding patterns of sequence-specific components of the general transcription machinery should be performed in ER-positive cells. This would be particularly interesting in MCF7 cells, in which, as suggested by the allele-specific differences in mRNA transcription observed in this cell line, those regulatory elements should be in an active state. Those functional elements might denote putative transcriptional enhancer or promoter regions in which a true functional SNP might reside (Chapter 3.4; (Birney, Stamatoyannopoulos et al. 2007; Koch, Andrews et al. 2007; Rada-Iglesias, Bajpai et al. 2011)). In addition, more exhaustive searches for other genetic variants closely linked to those SNPs, but not included in the HapMap and 1000 Genomes Project, might be necessary to develop a more comprehensive list of candidate functional SNPs.

3.5.3.2. A selected haplotype in the MDM4 gene regulates MDM4 mRNA expression

In the previous Chapter 3.4.2.2.1, we have shown that neutral haplotypes in the MDM4 gene, which associate with an increased risk of breast and ovarian cancers (Atwal, Kirchhoff et al. 2009; Kulkarni, Vazquez et al. 2009), also associate with

significantly higher MDM4 mRNA expression in cells. Hereby, NCI60 cell lines homozygous for the minor alleles expressed on average 2.6-fold more MDM4 mRNA than those homozygous for the major alleles (Chapter 3.4.2.2.1, Figure 3.4.6). In this Chapter, we have measured 1.4-fold more MDM4 RNA with the minor alleles of the haplotype in ERα-negative U2OS cells (Figure 3.5.1D), in line with the observations in the NCI60 panel and in support of the hypothesis that one or more functional SNP(s) from the MDM4 haplotype regulate the expression of MDM4 transcripts in an allele-specific manner. Interestingly, the polymorphic variants in the MDM4 gene are thought to associate with an earlier age of diagnosis of ER-negative, but not ER-positive, breast cancers (Kulkarni, Vazquez et al. 2009)(Gareth Bond, unpublished results, personal communication). Importantly, no allelic imbalance was observed in the ERα-positive MCF7 cell line (Figure 3.5.1D). Taken together, these results are potentially in line with the observation that the reported allelic differences in cancer diagnosis associated with the MDM4 SNPs depend on ER expression of the tumors. Possibly, the minor allele of an MDM4 SNP up-regulates the expression of full-length MDM4 mRNA in the absence of ERα signaling, whereas in the presence of ERα the relative up-regulation of full-length MDM4 transcription from the minor allele is reduced through transcriptional repression or mechanisms such as alternative splicing. Indeed, MDM4 is known to be regulated by alternative splicing and 6 different splice variants have been reported in the literature to date (Mancini, Di Conza et al. 2009). In one of best studied of these variants, termed MDM4-S, a premature stop codon in exon 7 results in a truncated form of the protein, which lacks the 3'UTR, a region to which our allele-specific primers/probes align (Rallapalli, Strachan et al. 1999; Bartel, Schulz et al. 2005). Therefore, a relative

induction of the expression of the MDM4-S variant in an ERα-positive setting would result in no measurable allelic imbalance with primers/probes designed to align to the 3'UTR region, in line with the observations in the MCF7 cells.

Together, these data support the hypothesis that one or more functional SNP(s) from the MDM4 haplotype regulate the expression of MDM4 transcripts in an allele-specific and ERα-dependent fashion. Others in our laboratory (Anna Grawenda, Alexander Finlayson) are currently utilizing this model to elucidate the exact functional mechanism behind the observed clinical and cellular phenotypes.

3.6. SNPs in the p53 pathway associate with increased risk for and worse prognosis of pancreatic ductal adenocarcinoma

3.6.1. Introduction

3.6.2 Results

3.6.2.1. MDM2 SNP309 associates with accelerated pancreatic ductal adenocarcinoma (PDAC) formation

3.6.2.2. CD44 and YWHAQ SNPs associate with allelic differences in PDAC survival and/or tumor onset

3.6.3. Discussion

3.6.3.1. The G-allele of MDM2 SNP309 could increase susceptibility to PDAC in a gender-dependent manner.

3.6.3.2. CD44 and YWHAQ SNPs could identify PDAC patients with a poor prognosis or individuals at higher risk of early tumor development

3.6.1. Introduction

Pancreatic ductal adenocarcinoma (PDAC) is the most frequent type of pancreatic cancer and accounts for eighty-five percent of the tumors originating in the pancreas (Garcea, Neal et al. 2005). It is an almost uniformly lethal disease due to its aggressive biological phenotype with an early local invasion and high metastatic potential (Yeo, Hruban et al. 2002; Hidalgo 2010). Only 20 percent of PDACs are amenable to surgical resection at presentation (Yeo, Hruban et al. 2002; Hidalgo 2010). Furthermore, the majority of cases presents in later stages of disease and are highly resistant to radiation- and chemotherapy, thereby resulting in very limited therapeutic options and an exceptionally poor overall prognosis, with a mean 5-year survival rate of below 5% (Yeo, Hruban et al. 2002; Hidalgo 2010).

The inactivation of the p53 pathway in the development of PDAC is a frequent event, whereby the p53 gene is mutated in 40-75% of tumors (Scarpa, Capelli et al. 1993; Pellegata, Sessa et al. 1994; Dong, Nio et al. 2003). In mouse models, p53 mutation has been shown to promote the development of invasive and metastatic PDAC in the context of Kras activation, the main oncogenic event in human PDAC (Yeo, Hruban et al. 2002; Hingorani, Wang et al. 2005; Bardeesy, Aguirre et al. 2006). A growing body of research suggests that inherited genetics can play a significant role in increasing the risk for PDAC (Yeo, Hruban et al. 2002; Hidalgo 2010). At least 6 cancer predisposition syndromes have been described that include an increased risk for PDAC development (Yeo, Hruban et al. 2002; Hidalgo 2010). Those predisposition syndromes include the Hereditary breast and ovarian cancer syndrome, the Familial atypical multiple mole melanoma and Peutz-

Jeghers syndromes (Giardiello, Brensinger et al. 2000; Murphy, Brune et al. 2002; Lynch, Fusaro et al. 2008; Hidalgo 2010). Individuals with these inherited disorders carry mutations in the BRCA1 and 2, CDKN2A and LKB1 genes, respectively (Giardiello, Brensinger et al. 2000; Murphy, Brune et al. 2002; Yeo, Hruban et al. 2002; Lynch, Fusaro et al. 2008; Hidalgo 2010). However, it has been estimated that these syndromes account for less than 10% of the hereditary pancreatic cancers and that there is no relation to these high-risk inherited syndromes in most of the cases in which there is a familial aggregation of the disease (Yeo, Hruban et al. 2002; Hidalgo 2010). The missing heritability could be explained by other, not yet identified low frequency, highly penetrant mutations, or multiple higher frequency, lower penetrant alleles of polymorphic loci. One example of such a locus could be the p53 network single nucleotide polymorphism in the MDM2 oncogene, MDM2 SNP309. Specifically, recent reports demonstrated that G-allele carriers of MDM2 SNP309 had an increased risk of pancreatic cancer and a decreased progression-free and overall survival in Northern American cohorts (Asomaning, Reid et al. 2008; Chen, Li et al. 2008).

In this study, we are interested in exploring the possibility that SNPs in p53 network genes might help identify a population at higher risk of PDAC or worse prognosis of this dismal disease, which could lead to the development of novel screening and treatment strategies, as well as a better understanding of the biology of this tumor. To this end, we first determined the genotypes of the most well described p53 network polymorphism, MDM2 SNP309, in a cohort of 107 German patients who underwent primary surgery for PDAC and 499 German blood donors. We attempt to validate the above-described associations of MDM2 SNP309 with PDAC risk and prognosis in this

cohort, as well as explore associations of this locus with both gender-and age-dependent PDAC incidence and p53 mutational status. We go on to explore potential associations of other functional p53 pathway SNPs, identified in this body of work, with PDAC risk, age-dependent incidence and disease outcome. The work presented in section 3.6.2.1 has been published during the continued work on my Thesis (Grochola, Muller et al. 2009).

3.6.2. Results

3.6.2.1. MDM2 SNP309 associates with accelerated pancreatic ductal adenocarcinoma (PDAC) formation

The MDM2 SNP309 (T/G) genotypes were determined successfully in 103 out of 107 PDAC patients. Thirty-five individuals (34%) were homozygous for the T-allele, 15 (14.6%) were homozygous for the G-allele and 53 patients (51.5%) were heterozygous (Table 3.6.1). In contrast to the previously published study in a cohort of Northern Americans from the Massachusetts General Hospital with a similar number of patients (n=109) (Asomaning et al 2008), no significant differences were found between the genotype frequencies in the patients and those of 499 blood donors of the same ethnic background (Table 3.6.1).

The MDM2 SNP309 locus has not only been shown to associate with an increased risk of some cancers, but also with an earlier age of diagnosis of cancer (Bond, Hu et al. 2004; Bond, Hirshfield et al. 2006; Bond and Levine 2007). Interestingly, in this PDAC cohort, patients G/G in genotype were diagnosed on average 9 years earlier in life (54.5 years, ranging from 32 to 72) than those harboring the T-allele (63.5 years, ranging

from 35 to 82, p=0.0210, T-Test, Figure 3.6.1A, Table 3.6.2). In contrast to studies of other cancers (Bond, Hirshfield et al. 2006; Bond and Levine 2007), the

Table 3.6.1. SNP309 allele frequency distribution: clinical and histopathological data. Data are number of patients.

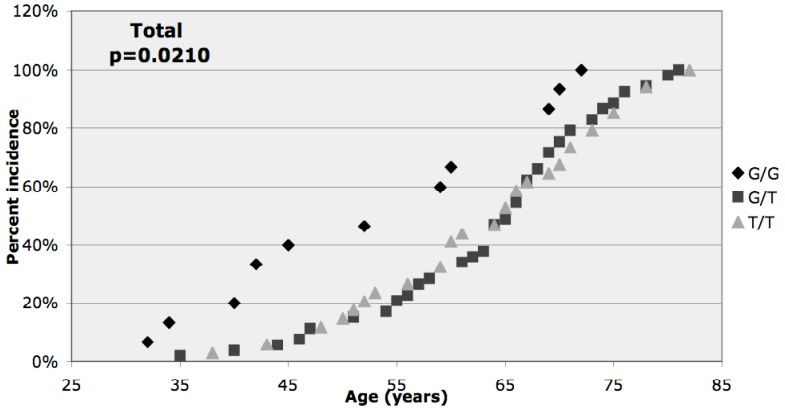
	Total	SNP309		
	n	T/T	T/G	G/G
Frequency (%)				
Controls	499	185 (37.1)	234 (46.9)	80 (16.0)
Cases	103	35 (34.0)	53 (51.5)	15 (14.6)
Stage (%)				
I	5	3 (60.0)	1 (20.0)	1 (20.0)
II	20	6 (30.0)	11 (55.0)	3 (15.0)
III	66	21 (31.8)	36 (54.5)	9 (13.6)
IV	12	5 (41.7)	5 (41.7)	2 (16.7)
Tumour resection (%)				
Radical (R0)	74	25 (33.8)	39 (52.7)	10 (13.5)
Not radical (R1)	18	6 (33.3)	9 (50.0)	3 (16.7)
Not radical (R2)	11	4 (36.4)	5 (45.5)	2 (18.2)

association of the G-allele with an earlier age of diagnosis was more pronounced in the male patients (Figure 3.6.1B and Table 3.6.2). Specifically, males G/G in genotype were diagnosed 12 years earlier (52.3 years) than males with the T-allele (64.3 years, p=0.0032, T-Test, Figure 3.6.1B and Table 3.6.2). A similar, but weaker, non-significant trend was observed in the female PDAC patients, whereby the females G/G in genotype were diagnosed 3.4 years earlier (59 years) than females with T-allele (62.4 years, p=0.5635, Figure 1C and Table 3.6.2).

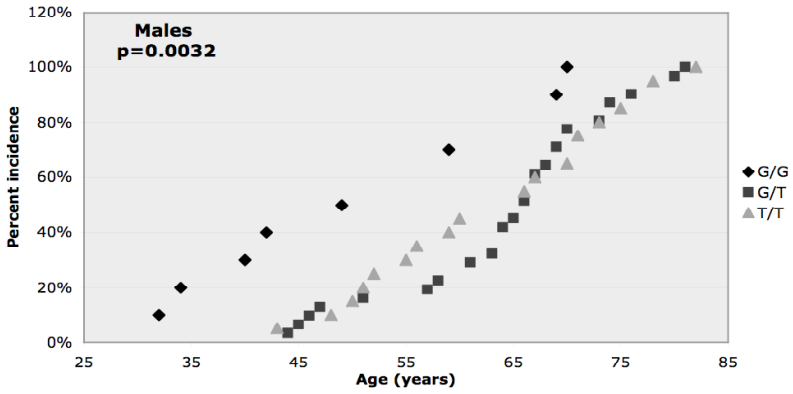
Table 3.6.2 Age of diagnosis. Data are given in years; * T-Test, T-allele vs G/G genotype.

	Total	SNP309			P-value*
	n	TT	TG	GG	
Cases					
Total	103	63.5	63.5	54.5	0.0210
Males	61	63.6	64.7	52.3	0.0286
Females	42	63.3	61.8	59.0	p>0.05

A



B



C

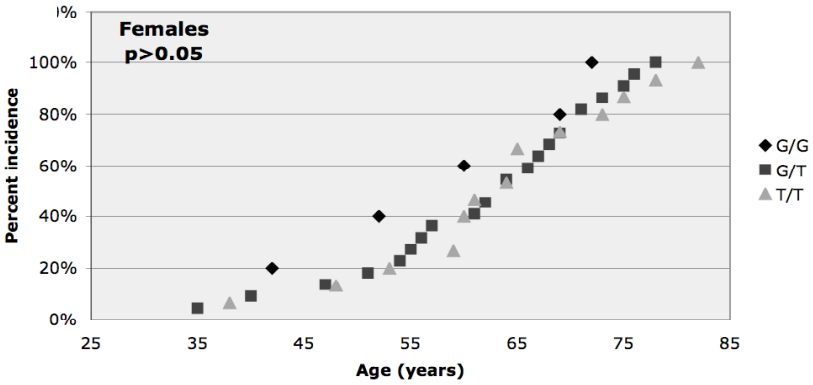


Figure 3.6.1. MDM2 SNP 309 allele expression: age of PDAC diagnosis. (A) 9-year earlier age of tumor diagnosis in patients homozygous for the G-allele compared to T-allele carriers, observed in 103 patients. (B) The results were more pronounced in males, whereby male patients with G/G genotype presented with a 12-year earlier tumor formation. (C) Same, however not significant, trend was seen in females. P-values were calculated using the T-Test.

In order to determine if the significant earlier age of diagnosis of PDAC in the male patients could suggest an increased risk of early onset disease, genotype frequencies were compared between patients diagnosed at younger ages and a control population. Male patients diagnosed at a younger age (below 51 years of age) were significantly enriched for the G-allele compared to older male patients and blood donors of the same ethnic background (Table 3.6.3). Specifically, men T/T in genotype make up only 18% of the male cases diagnosed at a younger age, as opposed to 36% of older male cases and 36% of younger male blood donors. In contrast, men G/G in genotype make up 45.5% of the male cases diagnosed at a younger age, as opposed to 10% of older male cases (OR=9; CI: 1.324-61.167; p=0.0256, Fisher's Exact Test, Table 3.6.3) and 15% of blood donors (OR=5.776; CI: 1.058-31.526; p=0.0379, Fisher's Exact Test, Table 3.6.3). No significant differences between the MDM2 SNP309 genotypes were observed between either the younger and older female patients or compared to the blood donors (Table 3.6.3). These data suggest that men G/G in genotype could have an increased risk for early onset PDAC with an odds ratio of 5.776 compared to an age- and gender-matched control population.

In contrast to previous reports, no significant associations were seen with the MDM2 SNP309 locus and overall survival. To assess the dependency of MDM2 SNP309 genotype on survival of this patient population, the impact on overall survival was first assessed by Kaplan-Meier analysis (Log-Rank test) and no significant differences were noted (Table 3.6.4). Specifically, patients with a G/G, T/G and T/T genotype had an estimated survival time of 20.4, 21.3 and 16.3 months, respectively

(p=0.467, log-rank test, Table 3.6.4). As the allelic differences in the age of PDAC diagnosis were gender-dependent, we were next interested in exploring

Table 3.6.3. Enrichments of G-allele carriers compared to blood donors suggest an increased risk of PDAC development only for younger males.

	MDM2 SNP309		Total	Females		Total	Males	
				< 51 yrs	≥ 51 yrs		< 51 yrs	≥ 51 yrs
Blood Donors	Allele (%)							
	T/T	185 (37.1)	72 (36.9)	57 (37.7)	15 (34.1)	113 (37.2)	67 (35.6)	46 (39.7)
	T/G	234 (46.9)	93 (47.7)	75 (49.7)	18 (40.9)	141 (46.4)	92 (48.9)	49 (42.2)
	G/G	80 (16.0)	30 (15.4)	19 (12.6)	11 (25.0)	50 (16.4)	29 (15.4)	21 (18.1)
	Total	499	195	151	44	304	188	116
PDAC Cases	Allele (%)							
	T/T	35 (34.0)	15 (35.7)	2 (40.0)	13 (35.1)	20 (32.8)	2 (18.2)	18 (36.0)
	T/G	53 (51.5)	22 (52.4)	3 (60.0)	19 (51.4)	31 (50.8)	4 (36.4)	27 (54.0)
	G/G	15 (14.6)	5 (11.9)	0 (0)	5 (13.5)	10 (16.4)	5 (45.5)	5 (10.0)
	Total	103	42	5	37	61	11	50
Case-Control Analyses	p-value Chi^2	0.8285	0.8450	0.6204	0.4365	0.6589	0.0282	0.7071
	p-value Fisher's Exact for enrichment of G/G*	1	0.7920	1	0.3607	0.8309	0.0379	0.4364
	Odds ratio for enrichement of G/G* (CI)	0.9911 (0.5125-1.916)	0.8000 (0.2667-2.399)	0.5897 (0.02710-12.835)	0.5245 (0.1440-1.910)	1.1300 (0.4932-2.589)	5.7760 (1.058-31.526)	0.6085 (0.1990-1.860)

* Fisher's Exact Tests compared to T/T carriers; Abbreviations. CI, Confidence interval

whether we can observe any gender-specific associations of MDM2 SNP309 on survival. When separated upon sex, the differences between the genotype did not become significantly more pronounced (Table 3.6.4). A multivariate Cox regression analysis adjusted for known independent prognostic factors for PDAC (tumor stage and resection type) also revealed no significant impact of the SNP309 genotype on tumor-related death (Table 3.6.5). In particular, an insignificant trend was observed for a worse overall survival for patients harboring the T-allele of SNP309. Hereby, individuals T/T in genotype and heterozygous at the SNP309 locus associated with a 1.835-fold (p=0.119) and 1.495-fold (p=0.274) increased relative risk for tumor-related death, respectively, compared to patients homozygous for the G-allele (Table 3.6.5). No further stratification according to the gender has been performed due to an insufficient sample size for multivariate survival analyses.

Table 3.6.4. MDM2 SNP309 - Kaplan Meier survival analysis.

Patient Group	Genotype	n	Estimated mean survival time (months)	p-value*
All patients	G/G	15	20.4	
	G/T	53	21.3	0.467
	T/T	35	16.3	
Females	G/G	5	27.5	
	G/T	22	21.3	0.513
	T/T	15	16.3	
Males	G/G	10	16.4	
	G/T	31	21.2	0.702
	T/T	20	15.0	

Table 3.6.5. MDM2 SNP309 - Cox's multivariate regression survival analysis.

Genotype	n	Relative risk*	p-value**
G/G	15	1	-
G/T	52	1.50	0.274
T/T	35	1.84	0.119

* Relative risk of tumor-related death

**P-values were calculated using the Cox's multivariate regression analysis, adjusted to tumor stage and resection-type

As mentioned above, the mutation of the p53 gene has been reported to be a frequent event in PDAC (40-75%) (Scarpa, Capelli et al. 1993; Pellegata, Sessa et al. 1994; Redston, Caldas et al. 1994; Dong, Nio et al. 2003). We determined the p53 mutational status in 75 of the patients' tumors by direct gene sequencing or microarray analysis. Interestingly, only fifteen (20%) were found to have p53 mutations, while the remaining 60 patients (80%) have retained a wild-type p53 gene. These mutations included 8 missense (10.7%) and 9 silent point mutations (12.0%, Materials and Methods Table 2.13, Chapter 2.2.4). No significant association(s) with p53 mutational status and the MDM2 SNP309 genotype was determined. Furthermore, no significant differences in genotype frequencies were seen for the histopathological and clinical parameters tumor stage and resection type (Table 3.6.1).

3.6.2.2. CD44 and YWHAQ SNPs associate with allelic differences in PDAC survival and/or tumor onset

The observation that MDM2 SNP309 associates with allelic differences in the development of pancreatic cancer supports a model that SNPs in the p53 pathway affect PDAC. In order to determine whether other inherited genetic variants in p53 network genes can affect the formation and/or the progression of PDAC, we first explore the effects of the two SNPs in the p53 network genes, CD44 and YWHAQ that were identified in Chapter 3.2 of this Thesis. These loci were shown to associate with allelic differences in cellular chemosensitivities, overall survival and age of soft-tissue sarcoma diagnosis. Of the 101 PDAC patients successfully genotyped for CD44 SNP, 43 (A/A,

42.5%) were homozygous for the A-allele, 11 (G/G, 10.9%) were homozygous for the G-allele and 47 (A/G, 46.6%) were heterozygous (Table 3.6.6). Thirty-three of the 104 patients (31.7%) successfully genotyped for the YWHAQ SNP were homozygous for the YWHAQ A-allele, 18 (17.3%) were homozygous for the G-allele and 53 patients (51%) were heterozygous (Table 3.6.6).

Table 3.6.6. Genotype frequencies CD44, YWHAQ and PPP2R5E SNP in PDAC

SNP	Genotype	Controls n (%)	Cases n (%)
CD44 SNP	TT	206 (41.4)	43 (42.5)
	CT	216 (43.4)	47 (46.6)
	CC	76 (15.3)	11 (10.9)
YWHAQ SNP	AA	139 (23.9)	33 (31.7)
	AG	242 (48.6)	53 (51.0)
	GG	117 (23.5)	18 (17.3)
PPP2R5E SNP	CC	266 (53.5)	47 (46.5)
	CT	193 (38.9)	51 (50.5)
	TT	38 (7.6)	3 (3.0)

We began by exploring potential associations of these loci with overall survival. To do this, we first performed a Kaplan-Meier analysis for the CD44 SNP. Interestingly, the analysis showed an insignificant trend, whereby PDAC patients harboring the allele of the CD44 SNP, which associated with poorer cellular chemosensitivities and prognosis in the STS patients, also associated with poorer overall survival in the PDAC cohort. Specifically, individuals with a C/C genotype had an estimated mean survival time of 11.64 months, whereas patients heterozygous and homozygous for the T-allele at the CD44 SNP locus associated with a longer overall survival of 20.01 and 20.02 months, respectively (p=0.141, Log-Rank Test; Figure 3.6.2A; Table 3.6.7). To exclude any

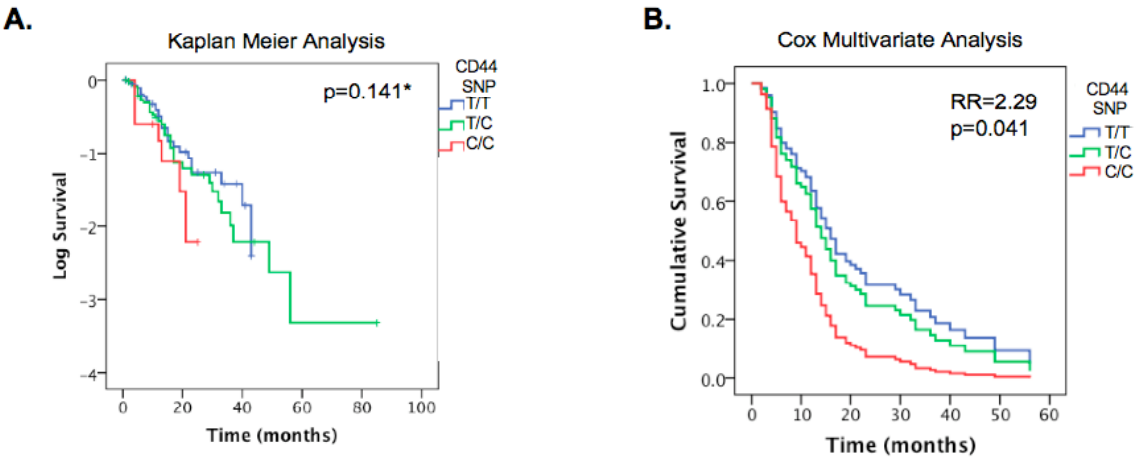


Figure 3.6.2. The alleles of the CD44 SNP, which associate with a weaker apoptotic response to chemotherapeutic treatment in the NCI60 cell lines, also correlate significantly with poorer overall survival in PDAC patients in a multivariate model. (A) Kaplan Meier survival estimates for CD44 SNP. (B) The graph displays the survival curves of all PDAC patients for the CD44 SNP genotypes plotted against the survival time in months. Patients with a C/C genotype associate with a 2.29-fold increased relative risk (RR) for tumor-related death compared to those T/T in genotype (Cox multivariate regression analysis, adjusted for the known prognostic factors of PDAC).

Table 3.6.7.Kaplan Meier survival estimates

CD44 SNP	n	Mean survival time (months)	p-value*
T/T	43	20.02	0.141
T/C	47	20.01	
C/C	11	11.64	

*log rank test T/T vs C/C

potential bias in the analysis due to independent prognostic factors, we next performed a multivariate Cox's regression analysis for the CD44 SNP adjusting for known independent prognostic factors for PDAC (tumor stage and resection type). Interestingly, and in line with our previous observations, PDAC patients harboring the C-allele of the CD44 SNP associated with a significantly poorer overall survival in the PDAC cohort.

Specifically, those individuals homozygous for the C-allele of the CD44 SNP associated with a 2.29-fold increased risk for tumor-related death compared to individuals T/T in genotype (p=0.041, Figure 3.6.2B, Table 3.6.8), while those heterozygous at this locus had an intermediate outcome. Despite the fact that the beneficial effects of chemotherapy are very limited in PDAC, some patients receive treatment with the nucleoside-analogue Gemcitabine, a standard first-line chemotherapeutic agent used for treatment of pancreatic cancer (Sultana, Tudur Smith et al. 2007; Hidalgo 2010). Only 52 out of the 101 individuals genotyped for CD44 SNP had received chemotherapy with Gemcitabine in our cohort, 45 had not undergone chemotherapy and the adjuvant treatment status of 4 patients was not known. Due to the low number of patients, who either had or had not received chemotherapy, we did not further stratify the patients according to the CD44 SNP genotypes in each group in order to perform a survival analysis. No significant allelic differences in overall survival were observed for the YWHAQ SNP in the Kaplan Meier and Cox multivariate analyses (Table 3.6.8), suggesting that the SNP does not affect the clinical outcome of PDAC in the same manner as STS.

Given our observations that MDM2 SNP309 associates with significant allelic differences in age of diagnosis in this PDAC cohort, we explored similar associations of the CD44 and YWHAQ loci. Moreover, we had previously seen allelic differences in the age of sarcoma diagnosis for both loci as we described in Chapter 3.2. Therefore, we explored the effects of these loci on the overall and age-dependent risk of PDAC. The genotype frequencies of 499 blood donors of the same ethnic background did not significantly differ to those found in the PDAC patients, suggesting that the SNPs

Table 3.6.8. Cox multivariate survival analysis

SNP	Patient group	Genotype	n	Relative risk*	p-value**
CD44 SNP	All Patients	T/T	42	1	-
		C/T	47	1.12	0.430
		C/C	11	2.29	0.041
	Patients with chemotherapy***	T/T	19	1	-
		C/T	27	1.031	0.932
		C/C	6	1.812	0.277
	Patients without chemotherapy***	C/C	21	1	-
		C/T	20	1.056	0.891
		T/T	4	4.322	0.095
YWHAQ (14-3-3 tau) SNP	All Patients	A/A	33	1	-
		A/G	52	1.11	0.694
		G/G	18	1.75	0.094
PPP2R5E SNP	All Patients	C/C	46	1	-
		C/T	51	1.51	0.090
		T/T	3	0.74	0.634

* Relative risk of tumor-related death

**P-values were calculated using the Cox's multivariate regression analysis, adjusted to tumor stage and resection-type

*** In 3 cases, it was not determined whether the patients received chemotherapy

do not associate with an overall increased risk for this cancer (Table 3.6.6), similar to our observations for these loci in the cohort of sarcoma patients (Chapter 3.2). Furthermore, no significant differences in genotype frequencies of the SNPs were seen for the histopathological and clinical parameters tumor stage and resection type (data not shown).

Next, we sought to determine whether the loci might affect the age-dependent incidence of this disease. Interestingly, patients harboring the A-allele of YWHAQ SNP were diagnosed with a mean age of 60.8 years (range 32-82 years), while patients G/G in genotype at 68.1 years (range 42-81 years). Thus, similar to our observations in sarcoma patients, PDAC patients carrying the A-allele showed on average a significant 7.3-year earlier age of onset than patients G/G in genotype ($p=0.0144$, Mann-Whitney Test, 2-tailed). Specifically, those individuals homozygous for the A-allele were diagnosed at

62.4 years of age (range 40-82), 5.7 years earlier than patients homozygous for the G-allele (p=0.0435, T-Test, one-sided, Table 3.6.9). In contrast to the observations in the soft-tissue sarcoma patients, no differences in the age of diagnosis between the genotypes of the CD44 SNP were found in the PDAC cohort (Table 3.6.9).

Table 3.6.9. Age of diagnosis analysis of CD44, YWHAQ and PPP2R5E SNPs in PDAC.

SNP	Genotype	n	Average Age of Diagnosis in Yrs (Median)	p-value (Mann-Whitney, T Test)*	Average Difference between Homozygouts
CD44 SNP	C/C	11	63.5 (64)	n.s.	1.8 years
	C/T	47	62.1 (65)		
	T/T	43	61.7 (64)		
YWHAQ (14-3-3 tau) SNP	G/G	18	68.1 (69.5)	0.0435**	5.7 years
	A/G	53	59.8 (61)		
	A/A	33	62.4 (64)		
PPP2R5E SNP	C/C	47	62.4 (65)	n.s.	6.7 years
	C/T	51	62.6 (64)		
	T/T	3	55.7 (54)		

*Major vs minor allele homozygouts; T-Test was chosen when data showed normal distribution, otherwise the Mann-Whitney Test was applied

** T-Test, 1-sided

In order to determine whether the observation that the A-allele of the YWHAQ SNP associates with a significantly earlier age of PDAC diagnosis could suggest an increased risk of early onset disease, genotype frequencies were compared between patients diagnosed at younger ages and a control population. Interestingly, patients diagnosed at a younger age (below 51 years of age) showed a trend for enrichment of the A-allele compared to older patients and blood donors of the same ethnic background (Table 3.6.10). Specifically, individuals carrying the A-allele make up 94.4% of the cases diagnosed at a younger age, as opposed to 80.2% of older cases (OR=4.188; CI: 0.5201-33.726; p=0.1881, Fisher's Exact Test, Table 3.6.10) and 76.3% of age- and gender-matched blood donors (OR=5.271; CI: 0.6904-40.250; p=0.0864, Fisher's Exact Test,

Table 3.6.10). However, these results were not significant and therefore failed to provide compelling support to the notion that the YWHAQ SNP associates with an increased risk for early onset disease.

Thus far, we have observed significant allelic differences for all p53 network SNPs in either age-dependent incidence or overall survival of PDAC. Therefore we next explored the ε-SNP2 in PPP2R5E that we identified in the screen for natural selection of p53 network genes and subsequently observed to associate with significant allelic

Table 3.6.10. Enrichments of G-allele carriers compared to blood donors suggest an increased risk of PDAC development only for younger males.

	YWHAQ SNP	Total	< 51 yrs	≥ 51 yrs
Blood Donors	Allele (%)			
	A-allele	381 (76.5)	258 (76.3)	123 (76.9)
	A/A	139 (27.9)	94 (27.8)	45 (28.1)
	A/G	242 (48.6)	164 (48.5)	78 (48.8)
	G/G	117 (23.5)	80 (23.7)	37 (23.1)
	Total	498	338	160
PDAC Cases	Allele (%)			
	A-allele	86 (82.7)	17 (94.4)	69 (80.2)
	A/A	33 (31.8)	7 (38.9)	26 (30.2)
	A/G	53 (51.0)	10 (55.6)	43 (50.0)
	G/G	18 (17.3)	1 (5.6)	17 (19.8)
	Total	104	18	86
Case-Control Analyses	p-value Chi^2 (AA-AG-GG)	0.1921	0.0905	0.5645
	p-value Fisher's Exact for enrichment of A-allele*	0.1966	0.0864	0.6289
	Odds ratio for enrichement of A-allele* (CI)	1.467 (0.8475-2.540)	5.271 (0.6904-40.250)	1.221 (0.6401-2.329)

* Fisher's Exact Tests compared to G/G carriers; Abbreviations. CI, Confidence interval

differences in the onset of, and risk for soft-tissue sarcoma development, as well as overall survival. Of the 101 patients successfully genotyped for ε-SNP2 in the German

PDAC patients and controls, 47 (C/C, 46.5%) were homozygous for the C-allele, 3 (T/T, 3%) were homozygous for the T-allele and 51 (C/T, 50.5%) were heterozygous (Table 3.6.6). These genotype frequencies did not significantly differ to those found in the blood donors, suggesting that the PPP2R5E SNP does not associate with an overall increased risk of this cancer. However, similar to the observations of this locus in the sarcoma cohort, the 3 PDAC patients T/T in genotype were diagnosed on average 6.9 years earlier in life at 55.7 years (ranging from 38 to 75, Table 3.6.9) than those heterozygous at this locus (mean of 62.6 years, ranging from 35 to 82) and 6.7 years earlier than patients homozygous for the C-allele who were diagnosed on average at 62.4 years (ranging from 32 to 82, p=0.2633, one-tailed Mann-Whitney Test, Table 3.6.9). However, the number of patients homozygous for the T-allele is very low in the PDAC cohort (n=3) and the results failed to reach statistical significance. No further separation based on gender was performed due to the low number of cases. In addition, no significant differences between the genotypes were observed with respect to postoperative survival (Table 3.6.8), association with tumor stage or resection-type of the tumors (data not shown). No further SNPs are genotyped in the PDAC cohort due to the unavailability of genomic DNA for further analyses.

3.6.3. Discussion

3.6.3.1. The G-allele of MDM2 SNP309 could increase susceptibility to PDAC in a gender-dependent manner.

The analyses of gender, age, and MDM2 SNP309 genotype in a cohort of 103 German PDAC patients suggest that the G-allele of MDM2 SNP309 accelerates tumor formation in PDAC, whereby patients G/G in genotype were diagnosed on average 9 years earlier than those carrying the T-allele (Figure 3.6.1A; Table 3.6.2; p=0.0271). Interestingly, further subset analysis revealed that this allelic difference is particularly pronounced in males, and not females, the former presenting with a significant 12-year earlier age of tumor formation (Figure 3.6.1B; Table 3.6.2; p=0.0215). Furthermore, in line with the gender-specific allelic differences in the age of PDAC diagnosis we observe that the G-allele of MDM2 SNP309 is significantly enriched in younger male patients compared to older male patients (Table 3.6.3; OR=9; p=0.0256) and an age- and gender-matched control population (Table 3.6.3; OR=5.776; p=0.0379). Taken together, these results suggest that the G-allele associates with an increased risk for early-onset PDAC disease. This observation requires further validation in other PDAC cohorts. Indeed, it is in stark contrast to previous reports in other tumor types that noted that the acceleration of tumorigenesis associated with the G-allele was dependent on the female gender and the primarily female specific hormone, estrogen. Specifically, the G-allele of MDM2 SNP309 was shown to accelerate tumorigenesis and increase cancer risk in women for colorectal cancer, diffuse large B-cell lymphoma, lung cancer, members of the Li-Fraumeni cancer predisposition syndrome with germline p53 mutations, highly estrogen

receptor positive (>50% of tumor cells), but not for estrogen receptor negative, invasive ductal carcinoma of the breast (Bond and Levine 2007; Atwal, Rabadan et al. 2008). This was shown to result in the enrichment of individuals with the G-allele in pre-menopausal women with these cancers, when compared to either post-menopausal women or men. Furthermore, evidence was provided for a possible molecular mechanism for how the G-allele could accelerate tumor formation in this gender-specific and estrogen-dependent manner (Hu, Feng et al. 2007). Hu et al. demonstrated that estrogen preferentially stimulated transcription of the MDM2 gene with the G-allele present compared to the T-allele possibly through the functional interactions between the estrogen receptor and its transcriptional co-activator SP1, which binds with greater affinity to the G-allele of the MDM2 promoter. Interestingly, SP1 is not only a defined co-activator for estrogen receptor, but also for the primarily male-specific androgen receptor (Lu, Jenster et al. 2000; Chen, Tsao et al. 2008). Furthermore, the presence of androgen receptors has been distinctly demonstrated in human PDAC tissues and human-derived pancreatic carcinoma cell lines (Corbishley, Iqbal et al. 1986; Konduri, Schwarz et al. 2007). It will be interesting to determine if male hormonal signaling could underlie this male-specific difference seen in this cohort.

Furthermore, our results are also different than those described in two recent reports, which utilized cohorts of equal or smaller sizes compared to our study, whereby a significant 2- to 3.6-fold increased risk for reduced overall- and/or progression-free survival for individuals with the G-allele of MDM2 SNP309 in PDAC was reported (Asomaning, Reid et al. 2008; Chen, Li et al. 2008). Interestingly, Chen et al. (Chen, Li et al. 2008) note that although no effect of MDM2 SNP309 on tumor progression in men

and women combined could be observed in their cohort comprising 88 patients, this polymorphism associated with a male-specific decrease of progression-free survival. We did not detect a decrease in survival for the G-allele of SNP309 in our in comparison larger cohort with uni- and multivariate analyses in either males or females. In contrast, we note an insignificant trend whereby patients harboring the G-allele associate with an improved overall survival (Table 3.6.4 and 3.6.5; up-to p=0.119, RR=1.84). These differences might be partly due to different treatments of the patients with DNA-damaging agents. A high percentage (92.7% and 100%, respectively) of patients received chemo-/radiotherapy in the above-mentioned studies compared to our cohort (52.5% of patients), which might imply different clinical responses with regard to MDM2 SNP309 and survival. Furthermore, none of the patients from our study received neoadjuvant radio chemotherapeutic treatment, while all patients in the study of Chen et al. (Chen, Li et al. 2008) were treated pre-operatively, and none in an adjuvant/palliative setting. However, we did not perform a further stratification of our cohort according to the status of adjuvant chemotherapeutic treatment due to an insufficient sample size for survival analyses, and therefore are not able to directly compare the results of both studies. In addition, Asomaning et al. reported a significant enrichment of G-allele carriers in their 109 PDAC patients comprising of Northern Americans when compared to a control population consisting of 372 individuals. These observations could also not be validated in our German cohort of a comparable study size (Table 3.6.3). Lastly, we have also incorporated the mutational status of p53 of the PDAC patients into our analysis of MDM2 SNP309, as MDM2 is a key negative regulator of p53 and its amplification in tumors has been suggested to alleviate the pressure for p53 mutations (Florenes,

Maelandsmo et al. 1994; Momand, Jung et al. 1998). Therefore, the regulation of MDM2 expression by MDM2 SNP309 might also affect the pressure on p53 to mutate in tumors and result in differences in genotype distributions in p53 wild-type and mutant tumors. However, we did not observe any pronounced or significant associations of MDM2 SNP309 genotype frequencies and the mutational status of p53, suggesting that MDM2 SNP309 does not significantly affect the requirement for p53 mutation in PDAC. Notably, the mutational status of p53 in our cohort is low in our cohort (20% frequency of p53 mutation), compared to reports from the literature, which note a frequency of p53 mutation ranging from 40% to 75% (Scarpa, Capelli et al. 1993; Pellegata, Sessa et al. 1994; Redston, Caldas et al. 1994; Dong, Nio et al. 2003). Possibly, these data might reflect a normal variation between the analyzed cohorts. However, most of the tumors from this study were analyzed by microarray analysis, in contrast to the use of sequencing and/or the single-strand conformation polymorphism (SSCP) method or PCR amplification with GC-clamped ampliprimers and subsequent denaturing gradient gel electrophoresis (DGGE) reported by others (Scarpa, Capelli et al. 1993; Pellegata, Sessa et al. 1994; Redston, Caldas et al. 1994). Therefore, the dissimilarities in the reported p53 mutation frequency might also be due to technical differences between the studies, and our report might understate the true p53 mutational frequency in our cohort.

In conclusion, the data provided here support a novel, male-predominant association of the G-allele of MDM2 SNP309 with a 12-year earlier age of tumor onset. Furthermore, a strong enrichment of the G-allele in younger males is observed, suggesting an increased risk for early onset PDAC with an odds ratio of up to 9. These

results lend further support to the notion that the G-allele can raise susceptibility to cancer in a gender-dependent manner.

3.6.3.2. CD44 and YWHAQ SNPs could identify PDAC patients with a poor prognosis or individuals at higher risk of early tumor development

In Chapter 3.2 of this Thesis, we report a screen of cellular chemosensitivity profiles that has led to the identification of polymorphisms in the p53 network genes 14-3-3 tau and CD44 that affect sarcoma incidence and survival. In this section, we aim to build on the findings from this screen and extend the analysis to patients diagnosed with pancreatic ductal adenocarcinomas. We genotyped the identified SNPs in the CD44 and YWHAQ genes in the well-described cohort of German PDAC patients. Interestingly, we find that, in line with the observations made in the initial screen (Chapter 3.2.2.2; Figure 3.2.4, Table 3.2.4), the allele of the CD44 SNP, which associates with weaker cellular growth responses to chemotherapeutics and poorer prognosis in the STS patients, also associates with worse overall survival in the PDAC cohort. Specifically, patients homozygous for the C-allele of CD44 SNP associate with a 2.29-fold increased relative risk for tumor-related death compared to those individuals T/T in genotype (Figure 3.6.2, p=0.041). No effect on patients' survival is observed for the YWHAQ SNP locus. In contrast, our analysis of these loci in the age-dependent incidence of PDAC revealed no significant differences for the CD44 SNP, but significant associations with the YWHAQ SNP. Specifically, patients carrying the A-allele of YWHAQ SNP show on average a 7.3-

year earlier age of onset than patients G/G in genotype (Table 3.6.9, p=0.0144), similar to the observations made in the sarcoma patients (Chapter 3.2.2.2, Table 3.2.5).

Besides its previously described roles in a vast range of cellular processes, such as regulation of growth and survival, differentiation and motility, CD44 is a key surface marker of cancer stem cells (CSC) in ductal adenocarcinoma of the pancreas, and believed to be involved in the mechanisms of self-renewal and multilineage differentiation as well as up-regulation of signaling cascades that are integral in tumor metastasis (Lee, Dosch et al. 2008). Importantly, the inhibition of pancreatic cancer stem cell characteristics has been recently shown to constitute an attractive therapeutic target and to have a promising potential for the future management of pancreatic cancer (Shankar, Nall et al. 2011). It is tempting to speculate that polymorphic variants of the CD44 gene might affect the activity or formation of CSCs in human pancreatic carcinoma and therefore affect the progression and prognosis of this cancer. In contrast to the observation made in the sarcoma cohort, we did not observe a significant association of the CD44 SNP with allelic differences in PDAC diagnosis, suggesting that CD44 SNP does not affect tumor initiation, as opposed to progression of PDAC. Certainly, the precise effects of altered CD44 activity on patients' risk and prognosis need to be further elucidated, but the reported studies and our data suggest a functionally and clinically important role of CD44 and its polymorphic variants in human pancreatic cancer.

The role of 14-3-3 proteins in PDAC has not been extensively studied yet, although the expression of the family member 14-3-3σ (sigma) has been shown to be elevated in human pancreatic adenocarcinoma (Logsdon, Simeone et al. 2003; Nakamura, Furukawa et al. 2004). Recently, it has been suggested that 14-3-3σ contributes to the

chemoresistance of pancreatic cancer cells and exerts cell type-dependent effects on cell migration and invasion (Neupane and Korc 2008). Our data suggest that 14-3-3τ might be involved in the formation of pancreatic cancer and its polymorphic variants might accelerate the initiation of carcinoma cells in this tumor type. This notion is consistent with the fact that 14-3-3τ has been shown to bind to wild-type p53 after cellular stresses and to stimulate transcriptional activity of p53 in cellular systems (Waterman, Stavridi et al. 1998; Stavridi, Chehab et al. 2001). Thus, polymorphic variants in the 14-3-3τ gene could affect the p53 stress response upon oncogenic activation and result in the observed allelic differences in the age of clinical tumor presentation. In contrast to the observation made in the sarcoma cohort, we did not observe a significant association of the YWHAQ SNP with allelic differences in PDAC prognosis, suggesting that YWHAQ SNP does not affect PDAC tumor progression. However, further studies aiming to analyze the precise role of YWHAQ in PDAC as well as to confirm the reported clinical associations of the YWHAQ SNP in this tumor are needed to elucidate the regulatory function of this SNP.

The role of the PPP2R5E gene in pancreatic cancer remains elusive, as no studies have reportedly analyzed its involvement in pancreatic cancer. We have observed an insignificant trend for the PPP2R5E ε-SNP2 locus, whereby in line with the observation in soft-tissue sarcomas, patients homozygous for the T-allele of PPP2R5E ε-SNP2 associate with a 6.7 years earlier age of PDAC diagnosis (Table 3.6.9). It is, of course, possible that the results failed to reach statistical significance because PPP2R5E does not affect PDAC formation and/or ε-SNP2 does not regulate PPP2R5E activity in this tissue type. However, it would be interesting to determine in larger pancreatic cancer cohorts, whether the same trend for earlier tumor onset is observed for PPP2R5E ε-SNP2 and the

lack of statistical significance is due to a small sample size in our cohort with only 3 patients carrying the T/T genotype.

In conclusion, the presented data support a model, whereby SNPs in p53 network genes have a potential to be utilized to predict patients at higher risk for tumor-related death and disease onset of PDAC. In line with the observations made in cellular chemosensitivity profiles and sarcomas, we show that the allele of the CD44 SNP, which is associated with weaker cellular growth responses to chemotherapeutics and poorer prognosis in the STS patients, also associates with worse overall survival in the PDAC cohort. Although no effect on patients' survival is observed for the YWHAQ SNP locus, this polymorphism associates with significant allelic differences in the age-dependent incidence of PDAC, again reflecting the observation made in the initial screen. The lack of significant associations in age-dependent incidence and overall survival for the CD44 and YWHAQ SNP locus, respectively, as well as for PPP2R5E ε-SNP2 might reflect tumor type-specific roles of the genes in tumor formation or progression.

IV. General Discussion

The p53 pathway constitutes a large cellular signaling network in vertebrates, which is crucial in directing the suppression of cancer formation, mediating the response to commonly used cancer therapies as well as the regulation of germ line maintenance, fertility and reproduction (Hu, Feng et al. 2008; Lane and Levine 2010). It has been demonstrated that low-frequency, highly penetrant, inherited mutations in p53 network genes underlie various cancer predisposition syndromes, such as the Li-Fraumeni Syndrome (Varley, Evans et al. 1997), Ataxia telangiectasia (Barlow, Hirotsune et al. 1996) or Cowden disease (Liaw, Marsh et al. 1997; Marsh, Dahia et al. 1997). Importantly, mounting evidence suggests that the knowledge of the germline genetics predisposing to those inherited cancer predisposition syndromes can have direct clinical implications and can be used to tailor therapies in a more personalized approach and improve survival (Hartmann, Schaid et al. 1999; Thull and Vogel 2004; Achatz, Hainaut et al. 2009; Link, Schuettpelz et al. 2011; Villani, Tabori et al. 2011).

The results from mouse models suggest that less penetrant alleles of p53 network genes that can be found at high frequencies in populations could also significantly alter p53 signaling in cells and affect cancer onset and progression (Mendrysa, McElwee et al. 2003; Bond, Hu et al. 2005; Alimonti, Carracedo et al. 2010). For example, mouse models designed to express hypomorphic alleles of MDM2, the key regulator of p53, have demonstrated that a slight reduction of MDM2 expression significantly affects p53 transcriptional activation and apoptotic activities, as well as tumor formation (Mendrysa, McElwee et al. 2003). Indeed, extensive studies of human genetic variants in p53 network genes have led to the identification of SNPs that result in allelic differences in

cancer risk, prognosis, responses to chemotherapy and show signs of being under evolutionary selective pressure (Grochola, Zeron-Medina et al. ; Atwal, Bond et al. 2007; Vazquez, Bond et al. 2008; Atwal, Kirchhoff et al. 2009). Detailed molecular and functional analyses of two SNPs in the p53 pathway that are most frequently studied (p53 codon72 (Harris, Brill et al. 1986) and MDM2 SNP309 (Bond, Hu et al. 2004)) suggest that the inherited genetics of high frequency genetic variation in the p53 network could contribute both to the understanding of the role of the host genes in human cancer, and to the improvement of current and future prevention and treatment strategies.

The aim of this project has been to further elucidate the role of SNPs in the p53 network of genes in cancer formation and progression. To achieve this, we first developed and applied three different screens to identify functional SNPs using characteristics of MDM2 SNP309 (Chapters 3.1-3.3). The screens searched for SNPs that associate with allelic differences in (i) recent natural selection, (ii) chemosensitivity profiles, and (iii) the gender- and age-dependent incidence of soft-tissue sarcoma. Using these screens, we identified SNPs in the PPP2R5E, CD44, YWHAQ, and ESR1 genes that associate with altered tumor formation, progression and response to chemotherapeutic treatment in cancer patients and cell lines. Secondly, we found epidemiological evidence in three independent cancer cohorts that the identified SNPs in the ESR1 gene interact with the known functional p53 polymorphism MDM2 SNP309 to increase cancer risk in an age- and gender-dependent manner.

In order to better understand the above-described cellular and clinical associations of the identified SNPs we are interested in the elucidation of the molecular and functional mechanisms underlying the described phenotypes. Indeed, a detailed molecular and

functional characterization of functional SNPs is crucial to deepen our understanding of inherited cancer susceptibility and outcome and to enable a more precise identification of individuals who carry a higher risk for cancer or have an altered ability to respond to cancer treatment. Therefore, in Chapter 3.4, we began to identify and characterize the causal SNP(s) in the haplotypes of the PPP2R5E, YWHAQ, CD44, and ESR1 genes that associate with the genomic, cellular and clinical phenotypes, as well as in the selected MDM4 haplotype, which was recently identified by our group (Atwal, Kirchhoff et al. 2009). We identified SNPs, which are in high linkage disequilibrium with each of the identified polymorphisms, and went on to identify those SNPs that reside in potential regulatory regions and could alter the expression of the respective genes. Interestingly, utilizing the NCI60 cell line panel and an isogenic cellular system, in which we could study the allelic imbalance in RNA levels in a single cell line, we observed allelic differences in RNA levels associated with SNPs in the MDM4, CD44, PPP2R5E and ESR1 genes. These data support the model, whereby those SNPs can affect the expression of the genes in which they reside, and thus possibly affect their protein levels and result in altered cellular activity, leading to an attenuation of the p53 pathway. The observed significant differences in RNA levels both in the cell line panel and in the heterozygous cells also suggests that these model systems can be successfully utilized to study the function of SNPs and their regulatory effects on the p53 pathway. They could serve as valuable experimental systems to test the predictions made based upon clinical and epidemiological observations as well as *in silica* analyses. Indeed, we attempted to utilize those cellular systems to study the molecular and functional mechanisms behind the clinical and cellular associations of the linked ESR1 SNPs in Chapter 3.5, which

provided additional experimental evidence that one or more SNPs could regulate ESR1 activity through affecting the expression of the ESR1 gene.

The first of the three screens aimed to identify functional p53 network SNPs (Chapter 3.1) evolved from the observation that haplotypes in the MDM2 and MDM4 genes, which associate with cancer-related phenotypes, deviated significantly from the standard assumptions of models of selective neutrality (Atwal, Bond et al. 2007; Atwal, Kirchhoff et al. 2009). We reasoned that signatures of natural selection could be used to identify other genes in the p53 pathway that harbor SNPs, which affect p53 signaling and human cancer. We performed a genomic scan for recent natural selection, which showed that of the 142 genes studied, the PPP2R5E gene that encodes a regulatory subunit of the tumor suppressing protein phosphatase 2A resides in a naturally selected genomic region. Subsequently, we demonstrated that a selected SNP in PPP2R5E (ε-SNP2) associates with significant allelic differences in the onset (up to 19.2 years; p=0.0002) and risk (odds ratio, up to 8.1; p=0.0009) of STS development, as well as overall survival (RR up-to 2.68; p=0.045). Together, these data added further support to the idea that signs of natural selection of p53 variants seem to be a characteristic of functional p53 pathway SNPs, which can be used to identify other p53 network genes, like PPP2R5E, that harbor SNPs that affect human cancer. Indeed multiple studies have lent support to this hypothesis and have identified candidate functional polymorphic loci in PPP2R2B (Vazquez, Kulkarni et al. 2011), TSC1 (Mehta, Vazquez et al. 2011), TSC2 (Mehta, Vazquez et al. 2011), and HAUSP (Kang, Feng et al. 2009) genes that harbor signs of evolutionary selection. Importantly, the reported observations remain to be validated in other patient cohorts, and the regulatory changes associated with PPP2R5E ε-SNP2 are

yet to be determined. However, the analysis described in Chapter 3.4 suggests a possible functional mechanism by which one or more SNPs from the selected haplotype could affect the described cancer phenotypes. Specifically, the SNPs that associated with the strongest signs of natural selection with derived (nonancestral) alleles are in high linkage disequilibrium with each other and reside in either intron 2 or intron 3, which could potentially function as distal transcriptional enhancers. Interestingly, we have found that cells homozygous for the minor allele of ε-SNP2, which associates with an earlier age of sarcoma diagnosis and increased overall risk for sarcoma in younger patients, expressed on average significantly, 1.47-fold higher PPP2R5E transcript levels compared to those with a C/C genotype. These data suggest that the haplotype linked to PPP2R5E ε-SNP2 affects the activity of PPP2R5E through an up-regulation of steady state mRNA levels. Interestingly, we observed that the allelic differences in mRNA expression are significantly more pronounced in cell lines, which express high levels of androgen receptor, suggesting that androgen regulation of PPP2R5E expression might account for the observed allelic differences in cancer risk, which are particularly pronounced in male patients. Further detailed molecular and functional analyses are needed to test those hypotheses and it will be interesting to determine the true functional SNP and the underlying regulatory mechanism(s). However, together these data suggest that PPP2R5E and the identified genetic variants play a significant regulatory role in the development and progression of soft-tissue sarcoma.

In the second screening approach (Chapter 3.2), we took advantage of a well-characterized trait of known-functional p53 SNPs to mine genomic datasets, namely their associations with altered cellular chemosensitivities (Vazquez, Bond et al. 2008), to

identify other functional SNPs in this important cancer-signaling pathway. A previous report had developed and implemented an analytic framework to uncover candidate functional SNPs in the p53 network utilizing data derived from the NCI60 human tumor cell line anticancer drug screen (Vazquez, Bond et al. 2008). This study identified 6 novel SNPs in 5 different genes, which showed significant allelic differences in growth response for the chemotherapeutic agents tested.

We expanded and utilized this methodology to scan 187 SNPs in 142 p53 stress response genes in order to identify potential functional p53 pathway SNPs. Hereby, we identified 7 SNPs in six genes that exhibit allelic differences in cellular growth-responses to many standard chemotherapeutic agents. Cells with different genotypes at these loci associated with, on average, a different response for 79 of the 132 agents tested, whereby the homozygouts differed on average 4.9-fold in their respective drug sensitivities. We observed the greatest differences in SNPs in the YWHAQ (14-3-3tau, rs6734469, $p=5.6x10^{-47}$) and CD44 (rs187115, $p=8.1x10^{-24}$) genes. Importantly, sarcoma patients with the alleles of these SNPs that associated with weaker cellular growth responses to chemotherapeutics, also associated with poorer overall survival (up-to 2.89 relative risk, p=0.011) and an earlier age of diagnosis (up-to 10.7 years earlier, p=0.002). Interestingly, preliminary data from our laboratory indicate that these alleles also associate with a significantly worse overall survival in patients diagnosed with chronic myelogenous leukemia (CML). Certainly, the observations reported in this Chapter remain to be validated in other patient cohorts and, importantly, the regulatory changes associated with YWHAQ and CD44 SNPs remain to be determined. However, the haplotype analysis in Chapter 3.4, together with the search for SNPs that reside in potential regulatory regions

of the genes, as well as our observations in the NCI60 cell line panel provide a hypothetical functional model for the identified SNPs in the CD44 gene, that can be tested in future experiments. Hereby, we observe that the CD44 allele, which associated with weaker responses to chemotherapeutic treatment, also showed a trend for an association with higher mRNA transcript levels in their respective genes, whereby a 4.35-fold higher mRNA transcript expression is observed in cells harboring the C-allele compared to those homozygous for the T-allele ($p=0.021$). We found that 6 out of 7 SNPs in intron 1 of the CD44 gene, which are linked to SNP rs187115, reside in potential genomic regulatory regions, suggesting that one of those SNPs might possibly reside in an enhancer or intronic promoter region and therefore affect the expression of the CD44 gene in an allele-specific manner. These results provide an experimentally verifiable model, whereby a minor allele of any of those closely linked polymorphisms in the CD44 gene, might lead through yet unidentified mechanisms to an elevated CD44 mRNA expression, possibly resulting in elevated CD44 protein levels and an increased oncogenic activity of this gene. Together, these data support the model that genetic variants in the CD44, and possibly the 14-3-3τ gene, play a significant regulatory role in cellular stress responses, thereby affecting sarcoma incidence and survival.

The third screening approach (Chapter 3.3.) evolved from the observation that the predominantly female G-allele carriers of MDM2 SNP309 are diagnosed with various cancers significantly earlier in life than T-allele carriers (Bond, Hirshfield et al. 2006; Bond, Menin et al. 2006; Bartel, Jung et al. 2008). Detailed molecular and functional studies provided convincing evidence that these gender-specific and estrogen receptor-dependent, clinical associations are the result of the transcriptional regulation of MDM2

gene expression by SP1 and its co-transcriptional factor estrogen receptor alpha (Bond, Hu et al. 2004; Hu, Feng et al. 2007). Together, these observations suggested that MDM2 SNP309 could change the effects of estrogen on cancer formation, leading to the observed gender- and age-dependent allelic differences in cancer incidence.

Therefore, we aimed to search for other SNPs in the p53 network that could potentially identify women, which are more susceptible for the tumor promoting effects of estrogen signaling. To this end, we chose the tumor type soft-tissue sarcoma in our screen, a p53-surveilled tumor, in which the gender-specific effects of MDM2 SNP309 were first described. Using a tag-SNP approach, we determined the genotypes of 639 SNP in 113 p53 network genes and identified a group of linked SNPs in the ERα gene that associates with altered age-dependent onset of STS solely in females. Specifically, we noted that females T/T in genotype for ESR1 $SNP^{rs2234693}$ were diagnosed on average 11.7-years earlier than those homozygous for the C-allele (p=0.0374). In addition, we demonstrated that the identified linked ESR1 SNPs significantly interact with MDM2 SNP309 to affect cancer formation in three independent cancer cohorts. Hereby, we observed that the allelic differences in the age of cancer diagnosis at the ESR1 $SNP^{rs2234693}$ locus are only observed in female, but not male, SNP309 G-allele carriers, while no effect is observed in females T/T for MDM2 SNP309. Specifically, in the German STS cohort, females carrying the G-allele of MDM2 SNP309, who are homozygous for the T-allele of ESR1 $SNP^{rs2234693}$, were diagnosed with STS on average 22.5 years (median 32.5 years) earlier than those carrying the G-allele of MDM2 SNP309 and C/C in genotype at the ESR1 $SNP^{rs2234693}$ locus (p=0.031). Importantly, these findings were further supported and validated by similar observations in a second

sarcoma cohort from Australia, whereby in females, who harbor the MDM2 SNP309 G-allele, patients carrying a T-allele of ESR1 SNP^{rs2234693} associate with a 16.1-year earlier age of tumor diagnosis compared to those C/C in genotype (p=0.0207). In addition, we observed the same significant trend in ovarian carcinoma patients who carry an MDM2 SNP309 G-allele and whose tumors express the estrogen receptor, with a 7.3-year allelic difference between females carrying the T-allele of ESR1 SNP^{rs2234693} and those C/C in genotype (p=0.0414). Together, these results support the hypothesis that the identified ESR1 SNPs interact with MDM2 SNP309 in a gender-specific and estrogen receptor-dependent manner to affect human cancer risk, whereby the effects of the ESR1 SNPs are only found in female G-allele carriers and/or tumors that express ER. This is consistent with the previously proposed model that ER is more effective at activating MDM2 transcription from the G-allele of MDM2 SNP309 (Hu, Feng et al. 2007)

Interestingly, our data, together with a previous report analyzing ESR1 mRNA expression in brain tissues (Weickert, Miranda-Angulo et al. 2008), suggest that the major allele(s) of one or more of the identified ESR1 SNPs function through an up-regulation of the expression of the gene. Specifically, we observe a higher ESR1 expression associated with the major T-allele of ESR1 SNPrs2234693, compared to the minor C-allele in the heterozygous MCF7 cell line, a breast cancer cell line panel selected based upon reported ER-positivity (Neve, Chin et al. 2006), as well as in the NCI60 cell line panel. We identified 16 SNPs in high linkage disequilibrium with the identified SNPs, one of which resides in a highly conserved genomic region in intron 1 of the ESR1 gene that could constitute a distal enhancer. We went on to demonstrate in Chapter 3.5 that this SNP significantly changes the *in vitro* binding affinity of transcription factors

from the bHLH family, which bind to E-Box elements. Interestingly we show in the heterozygous MCF7 cell line, as well as possibly in a breast cancer cell panel that the bHLH transcription factor AP4 regulates ESR1 mRNA expression in an allele-specific manner, in line with the *in vitro* observed results. However, the precise mechanisms of how AP4 regulates ESR1 expression remain elusive, as we have not been able to detect a binding of AP4 to the SNP sequence *in vivo* using a ChIP technique, nor have we been able to demonstrate an allele-specific enhancer function of this region in reporter assays. Therefore, the exact mechanism by which the identified ESR1 SNPs affect cancer risk still needs further investigation. Further molecular and functional analyses are necessary. First, the regulatory mechanisms associated with these SNPs need to be described in detail in order to extensively test the model, that a major allele of one (or more) SNP(s) changes the binding affinity of sequence-specific transcription factors and up-regulates the expression of the ESR1 gene. In this model, the resulting increased ERα protein levels could lead to a heightened proliferation in cells and an increased cancer risk. Once a precise functional mechanism has been established, it will be of great interest to determine how a functional SNP in the ESR1 gene interacts with MDM2 SNP309 to affect cancer risk. Based upon reports from the literature, one possible model would be that the hypothesized higher ERα expression associated with the major alleles of the ESR1 haplotype leads to an increased activation of the MDM2 promoter through SP1, a co-transcriptional factor of ERα, which has been shown to regulate MDM2 expression in an estrogen-dependent manner. Specifically, through the known higher binding affinity of the SP1-ERα complex to the G-allele of SNP309, the effect of elevated ERα levels on MDM2 transcription would be most pronounced in cells with the G-allele. This, in turn

could result in increased MDM2 protein levels, an attenuation of the tumor suppressive role of p53 and an increased tumor formation. It will also be interesting in the future to determine whether both SNPs can also interact to affect tumor progression or response to therapy, which might carry important clinical implications. For example, upon speculation, treatment with (selective) ER antagonists might become a therapeutic option worth considering in those cancer patients with a certain combination of genotypes, such as female individuals who carry a SNP309 G-allele and an ESR1 SNP^{rs2234693} T-allele, as the tumors of those patients might have developed an "addiction" for estrogen signaling.

The epidemiological evidence presented in this study for an interaction of an ESR1 SNP with MDM2 SNP309, together with the observation that other SNPs, like MDM2 SNP285 (Knappskog, Bjornslett et al. 2011) or p53 codon72 polymorphism (Wan, Wu et al. 2011) can significantly affect the associations of SNP309 with cancer risk, suggests that different alleles of p53 network genes can interact to further weaken the pathway and increase the susceptibility to cancer. The knowledge of such interactions could prove valuable to characterize individuals in the strength of the p53 pathway stress response and could therefore offer novel treatment and prevention strategies, as well as personalize existing ones. It will be necessary to identify more SNPs in this important tumor suppressor pathway and, crucially, characterize those SNPs on a functional and molecular level, in order to understand the specific biological context in which a SNP, alone or in interaction with other polymorphisms, can affect cancer. Hopefully, this approach, with the aim to gain a deep understanding of the molecular biology behind those clinical associations, will also help to reduce the number of false-positive associations of SNPs with cancer, as has been successfully achieved in genome-wide

association studies (GWAS) through the use of high statistical stringency (Chen and Thomas 2010). In addition, we hope that our approach using biological knowledge will also enable to discover truly causative SNPs, which are left undiscovered by the agnostic GWAS methodology (Grochola, Zeron-Medina et al. ; Chen and Thomas 2010), due to the dependence of those SNPs on specific biological contexts, such as gender, tissue-specificity or specific somatic mutations.

Lastly, in Chapter 3.6, we were interested in exploring the possibility that SNPs in p53 network genes might help identify a population at higher risk or worse prognosis of pancreatic ductal adenocarcinoma, a dismal disease due to its aggressive biological phenotype and a high resistance to conservative treatment (Yeo, Hruban et al. 2002; Hidalgo 2010). In order to help identify a population at higher risk or worse prognosis of PDAC, for which the genetic basis for the predisposition has yet to be determined, we first genotyped MDM2 SNP309 in a cohort of 107 German PDAC patients and 499 German blood donors. We observed that the G-allele of MDM2 SNP309 associated with a 9-year earlier age of PDAC diagnosis ($p=0.021$). Interestingly, however, in contrast to studies of other tumor types, these observations were made predominantly in men and not women, whereby conditions of male PDAC patients with a G/G genotype were diagnosed on average 12 years earlier than T-allele carriers ($p=0.0032$). Furthermore, particularly younger male patients presented a significant enrichment of the G-allele ($p=0.0282$), suggesting a novel role of the MDM2 SNP309 locus in regulating PDAC tumor formation in a male-specific manner, which merits further study.

In order to determine whether other SNPs in p53 pathway can affect tumorigenesis of PDAC or the progression of the disease, we next explored the effects of

the CD44, YWHAQ and PPP2R5E SNPs that were identified in this Thesis. Interestingly, we found that, in line with the observations made in Chapter 3.2, the allele of the CD44 SNP, which associated with weaker cellular growth responses to chemotherapeutics and poorer prognosis in the STS patients, also associated with an increased relative risk for tumor-related death in the PDAC cohort (RR=2.29, p=0.041), further supporting the hypothesis that CD44 SNP plays a significant regulatory role in cellular stress responses, thereby affecting cancer survival. Other associations of the identified SNPs observed in the sarcoma cohort were not repeated in PDAC. For example, no association with patients' survival was observed for the YWHAQ SNP locus and the minor allele of PPP2R5E ε-SNP2 was only found at a very low frequency in PDAC. There are many possible reasons for the differences between pancreatic carcinoma and sarcoma in the effects of the studied p53 pathway SNPs. For example, many effects of SNPs might be tissue-specific, and therefore significant associations in one tumor type might not be observed in others. Secondly, the patient cohorts utilized in our study might lack a sufficient sample size to detect small differences in cancer phenotypes. Although the reasons for these differences remain to be determined, and the observations made in our study need to be repeated in other pancreatic cancer cohorts, our results lend support to the notion that p53 network SNPs can affect susceptibility to PDAC and affect disease outcome. Potentially, these SNPs could contribute to the better understanding of the genetic basis of PDAC and help in the personalization of treatment strategies of this dismal disease.

Importantly, our studies also have limitations, which have to be considered in the careful interpretation of the results and the design of future studies. Firstly, we utilized

cohorts with relatively small sample sizes, which carry the risks of both false-positive and false-negative associations. An increase in the sample size would lower the probability that a certain genotype-phenotype constellation is found simply by chance (false-positive result) or that a smaller, but significant effect is left undiscovered due to an insufficient power to detect weaker associations (false-negative result). Furthermore, the utilization of only one cohort of a certain tumor type carries the risk that the observed effects are specific for the studied population or are caused by hidden biases in the recruitment of the cohort. To address this point, it will be important to validate our findings in similar but unrelated cohorts, a consideration which will be also important in the design of future studies. Furthermore, the cell culture based studies utilized in this Thesis also carry important limitations. We used the NCI60 cell line panel to search for allelic-differences in cellular chemosensitivities and gene expression levels. This panel, which comprises 59 cell lines, is composed of 9 different tumor types, and therefore tissue-specific effects might bias the analyses of SNPs. Larger cell line panels comprising many cell lines derived from the same tumor type(s), such as the tumor-derived cell line panel from the currently conducted Cancer Genome Project (CGP) (Stratton, Campbell et al. 2009) will be able to largely address those concerns. In addition, the ENCODE dataset we utilize to search for putative regulatory regions in Chapter 3.4 at the present stage incorporates data obtained from only a limited amount of cell types. Therefore, tissue-, cell- and developmental stage-specific effects might complicate the interpretation of the experimental data. Furthermore, the project covers only 1% of the human genome and therefore many regulatory regions in which SNPs reside might not be covered by the project. However, efforts are under way at present to significantly increase both the

variety and number of cell lines studied as well as to extend the genomic coverage of the project with the advent of high-throughput experimental techniques (Rosenbloom, Dreszer et al. 2010; Myers, Stamatoyannopoulos et al. 2011). Lastly, despite the continuous and extraordinary progress in the identification and cataloguing of SNPs in the past decades (Chapter 1.3.3), it is likely that new SNPs are yet to be discovered. The most comprehensive dataset of SNPs with information about their linkage can be found at present in the Pilot 1 of the 1000 Genomes Project, which utilized whole-genome sequencing technologies with an average sequencing depth of 3.6X to identify and map polymorphisms (Durbin, Abecasis et al. 2010). The improvement of the sequencing depth to above 4X upon completion of the project (Durbin, Abecasis et al. 2010), and possibly even higher in the more distant future, might lead to the annotation of new SNPs in the present databases.

Taken together, our work suggests that the inherited genetics of p53 pathway indeed have a potential to further define populations in their abilities to react to stress, suppress tumor formation and respond to therapies. Utilizing well-characterized traits of known-functional p53 SNPs, we successfully search for and identify novel SNPs in the CD44, YWHAQ, PPP2R5E and ESR1 genes that affect cancer risk, age-dependent incidence, response to chemotherapeutic treatment and patients' prognosis in sarcoma, ovarian carcinoma and pancreatic cancer. We hope that the framework designed in this Thesis to identify and characterize functional SNPs can be successfully implemented and optimized in future studies to improve our knowledge of human genetic variation in this important tumor suppressor pathway, with the ultimate goal to improve cancer prevention and treatment strategies.

VI. Acknowledgments

I would like to thank, first of all, my supervisor Dr. Gareth Bond for his exceptional and cordial supervision, instructive discussions and genuine support during all stages of my DPhil studies.

Prof. Peter Würl for the contribution of the German soft-tissue sarcoma and pancreatic carcinoma cohorts and control DNAs for this study and his valuable continued support.

Dr. Alexei Vazquez for the productive collaboration on the work presented in Chapters 3.1 and 3.2.

Prof. Thomas Möller for the contribution of the control DNAs.

Prof. Helge Taubert, for the contribution of the German soft-tissue sarcoma cohort.

Dr. David Thomas for the contribution of the Australian sarcoma cohort.

Dr. Frank Bartel for the contribution of the ovarian carcinoma cohort.

Prof. Xin Lu for the opportunity to study at the Ludwig Institute.

Dr. Sarah De Val, Heather Shaw, Mario Notari, Dr. John Christianson, Prof. Colin Goding and Dr. Eric O'Neill for their valuable help in the laboratory and insightful discussions.

I also thank Elisabeth Bond for her friendly and professional support in the lab. All other lab members, Jorge Zeron, Emmanouela Repapi and Anna Grawenda for their valuable help with this project and their cordial support.

The work conducted during the work on this Thesis and presented in Chapters 3.1, 3.2 and 3.6 have been conducted in collaboration with Dr. Alexei Vazquez from Princeton University, USA; Prof. Dr. Peter Würl from the University Clinic of Ulm, Germany and Prof. Dr. Thomas Müller from the German Red Cross Centre in Springe, Germany. All parts of my Thesis, which are not my own work are clearly indicated and referenced in the manuscript. The work has been subject to three publications, all of which I have co-authored as first or second major contributing author with shared first authorship (Grochola, Muller et al. 2009; Grochola, Vazquez et al. 2009; Vazquez, Grochola et al. 2010). In addition, the material presented in Chapter 1 has been published in part as a review article (Grochola, Zeron-Medina et al. 2010). I have written the article during the work on my Thesis under the supervision of Dr. Gareth Bond. Permission from the co-authors of all 4 publications has been granted to include the material in my Thesis.

VI. References

Abbas, T. and A. Dutta (2009). "p21 in cancer: intricate networks and multiple activities." Nat Rev Cancer **9**(6): 400-414.

Abukhdeir, A. M. and B. H. Park (2008). "P21 and p27: roles in carcinogenesis and drug resistance." Expert Rev Mol Med **10**: e19.

Achatz, M. I., P. Hainaut, et al. (2009). "Highly prevalent TP53 mutation predisposing to many cancers in the Brazilian population: a case for newborn screening?" Lancet Oncol **10**(9): 920-925.

Achatz, M. I., M. Olivier, et al. (2007). "The TP53 mutation, R337H, is associated with Li-Fraumeni and Li-Fraumeni-like syndromes in Brazilian families." Cancer Lett **245**(1-2): 96-102.

Alder, H., M. Yoshinouchi, et al. (1992). "A conserved region in intron 1 negatively regulates the expression of the PCNA gene." Nucleic Acids Res **20**(7): 1769-1775.

Alhopuro, P., S. K. Ylisaukko-Oja, et al. (2005). "The MDM2 promoter polymorphism SNP309T-->G and the risk of uterine leiomyosarcoma, colorectal cancer, and squamous cell carcinoma of the head and neck." J Med Genet **42**(9): 694-698.

Alimirah, F., J. Chen, et al. (2006). "DU-145 and PC-3 human prostate cancer cell lines express androgen receptor: implications for the androgen receptor functions and regulation." FEBS Lett **580**(9): 2294-2300.

Alimonti, A., A. Carracedo, et al. (2010). "Subtle variations in Pten dose determine cancer susceptibility." Nat Genet **42**(5): 454-458.

Altshuler, D. M., R. A. Gibbs, et al. (2010). "Integrating common and rare genetic variation in diverse human populations." Nature **467**(7311): 52-58.

American_College_of_Medical_Genetics_Board_of_Directors (1995). "Points to consider: ethical, legal, and psychosocial implications of genetic testing in children and adolescents. American Society of Human Genetics Board of Directors, American College of Medical Genetics Board of Directors." Am J Hum Genet **57**(5): 1233-1241.

Anghel, A., D. Narita, et al. (2010). "Estrogen Receptor Alpha Polymorphisms and the Risk of Malignancies." Pathol Oncol Res.

Antequera, F., J. Boyes, et al. (1990). "High levels of de novo methylation and altered chromatin structure at CpG islands in cell lines." Cell **62**(3): 503-514.

Arroyo, J. D. and W. C. Hahn (2005). "Involvement of PP2A in viral and cellular transformation." Oncogene **24**(52): 7746-7755.

Arva, N. C., T. R. Gopen, et al. (2005). "A chromatin-associated and transcriptionally inactive p53-Mdm2 complex occurs in mdm2 SNP309 homozygous cells." J Biol Chem **280**(29): 26776-26787.

Ashcroft, M., M. H. Kubbutat, et al. (1999). "Regulation of p53 function and stability by phosphorylation." Mol Cell Biol **19**(3): 1751-1758.

Asomaning, K., A. E. Reid, et al. (2008). "MDM2 Promoter Polymorphism and Pancreatic Cancer Risk and Prognosis." Clin Cancer Res **14**(12): 4010-4015.

Atwal, G. S., G. L. Bond, et al. (2007). "Haplotype structure and selection of the MDM2 oncogene in humans." Proc Natl Acad Sci U S A **104**(11): 4524-4529.

Atwal, G. S., T. Kirchhoff, et al. (2009). "Altered tumor formation and evolutionary selection of genetic variants in the human MDM4 oncogene." Proc Natl Acad Sci U S A **106**(25): 10236-10241.
Atwal, G. S., R. Rabadan, et al. (2008). "An information-theoretic analysis of genetics, gender and age in cancer patients." PLoS ONE **3**(4): e1951.
Bandele, O. J., X. Wang, et al. (2011). "Human single-nucleotide polymorphisms alter p53 sequence-specific binding at gene regulatory elements." Nucleic Acids Res **39**(1): 178-189.
Banin, S., L. Moyal, et al. (1998). "Enhanced phosphorylation of p53 by ATM in response to DNA damage." Science **281**(5383): 1674-1677.
Barbour, A. P., J. A. Reeder, et al. (2003). "Expression of the CD44v2-10 isoform confers a metastatic phenotype: importance of the heparan sulfate attachment site CD44v3." Cancer Res **63**(4): 887-892.
Bardeesy, N., A. J. Aguirre, et al. (2006). "Both p16(Ink4a) and the p19(Arf)-p53 pathway constrain progression of pancreatic adenocarcinoma in the mouse." Proc Natl Acad Sci U S A **103**(15): 5947-5952.
Barlow, C., S. Hirotsune, et al. (1996). "Atm-deficient mice: a paradigm of ataxia telangiectasia." Cell **86**(1): 159-171.
Bartek, J., J. Falck, et al. (2001). "CHK2 kinase--a busy messenger." Nat Rev Mol Cell Biol **2**(12): 877-886.
Bartel, F., J. Jung, et al. (2008). "Both germ line and somatic genetics of the p53 pathway affect ovarian cancer incidence and survival." Clin Cancer Res **14**(1): 89-96.
Bartel, F., J. Schulz, et al. (2005). "Significance of HDMX-S (or MDM4) mRNA splice variant overexpression and HDMX gene amplification on primary soft tissue sarcoma prognosis." Int J Cancer **117**(3): 469-475.
Beckerman, R. and C. Prives (2010). "Transcriptional regulation by p53." Cold Spring Harb Perspect Biol **2**(8): a000935.
Beckman, G., R. Birgander, et al. (1994). "Is p53 polymorphism maintained by natural selection?" Hum Hered **44**(5): 266-270.
Beckmann, M. W., D. Niederacher, et al. (1997). "Multistep carcinogenesis of breast cancer and tumour heterogeneity." J Mol Med **75**(6): 429-439.
Belyi, V. A., P. Ak, et al. (2010). "The origins and evolution of the p53 family of genes." Cold Spring Harb Perspect Biol **2**(6): a001198.
Benard, J., S. Douc-Rasy, et al. (2003). "TP53 family members and human cancers." Hum Mutat **21**(3): 182-191.
Bergamaschi, D., M. Gasco, et al. (2003). "p53 polymorphism influences response in cancer chemotherapy via modulation of p73-dependent apoptosis." Cancer Cell **3**(4): 387-402.
Bergamaschi, D., Y. Samuels, et al. (2006). "iASPP preferentially binds p53 proline-rich region and modulates apoptotic function of codon 72-polymorphic p53." Nat Genet **38**(10): 1133-1141.
Bethke, L., K. Sullivan, et al. (2008). "The common D302H variant of CASP8 is associated with risk of glioma." Cancer Epidemiol Biomarkers Prev **17**(4): 987-989.

Bharadwaj, R. R., C. D. Trainor, et al. (2003). "LCR-regulated transgene expression levels depend on the Oct-1 site in the AT-rich region of beta -globin intron-2." Blood **101**(4): 1603-1610.

Birgander, R., A. Sjalander, et al. (1996). "The codon 31 polymorphism of the p53-inducible gene p21 shows distinct differences between major ethnic groups." Hum Hered **46**(3): 148-154.

Birney, E., J. A. Stamatoyannopoulos, et al. (2007). "Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project." Nature **447**(7146): 799-816.

Bittenbring, J., F. Parisot, et al. (2008). "MDM2 gene SNP309 T/G and p53 gene SNP72 G/C do not influence diffuse large B-cell non-Hodgkin lymphoma onset or survival in central European Caucasians." BMC Cancer **8**: 116.

Blanco-Aparicio, C., O. Renner, et al. (2007). "PTEN, more than the AKT pathway." Carcinogenesis **28**(7): 1379-1386.

Blencowe, B. J. (2000). "Exonic splicing enhancers: mechanism of action, diversity and role in human genetic diseases." Trends Biochem Sci **25**(3): 106-110.

Bodmer, J. L., N. Holler, et al. (2000). "TRAIL receptor-2 signals apoptosis through FADD and caspase-8." Nat Cell Biol **2**(4): 241-243.

Bolderson, E., D. J. Richard, et al. (2009). "Recent advances in cancer therapy targeting proteins involved in DNA double-strand break repair." Clin Cancer Res **15**(20): 6314-6320.

Bonafe, M., S. Salvioli, et al. (2002). "p53 codon 72 genotype affects apoptosis by cytosine arabinoside in blood leukocytes." Biochem Biophys Res Commun **299**(4): 539-541.

Bond, G. L., K. M. Hirshfield, et al. (2006). "MDM2 SNP309 Accelerates Tumor Formation in a Gender-Specific and Hormone-Dependent Manner." Cancer Res **66**(10): 5104-5110.

Bond, G. L., W. Hu, et al. (2004). "A single nucleotide polymorphism in the MDM2 promoter attenuates the p53 tumor suppressor pathway and accelerates tumor formation in humans." Cell **119**(5): 591-602.

Bond, G. L., W. Hu, et al. (2005). "MDM2 is a central node in the p53 pathway: 12 years and counting." Curr Cancer Drug Targets **5**(1): 3-8.

Bond, G. L. and A. J. Levine (2007). "A single nucleotide polymorphism in the p53 pathway interacts with gender, environmental stresses and tumor genetics to influence cancer in humans." Oncogene **26**(9): 1317-1323.

Bond, G. L., C. Menin, et al. (2006). "MDM2 SNP309 accelerates colorectal tumour formation in women." J Med Genet **43**(12): 950-952.

Bornstein, P., J. McKay, et al. (1988). "Interactions between the promoter and first intron are involved in transcriptional control of alpha 1(I) collagen gene expression." Mol Cell Biol **8**(11): 4851-4857.

Bougeard, G., S. Baert-Desurmont, et al. (2006). "Impact of the MDM2 SNP309 and p53 Arg72Pro polymorphism on age of tumour onset in Li-Fraumeni syndrome." J Med Genet **43**(6): 531-533.

Brest, P., P. Lapaquette, et al. (2011). "A synonymous variant in IRGM alters a binding site for miR-196 and causes deregulation of IRGM-dependent xenophagy in Crohn's disease." Nat Genet **43**(3): 242-245.

Brooks, C. L. and W. Gu (2003). "Ubiquitination, phosphorylation and acetylation: the molecular basis for p53 regulation." Curr Opin Cell Biol **15**(2): 164-171.

Brown, J. P., W. Wei, et al. (1997). "Bypass of senescence after disruption of p21CIP1/WAF1 gene in normal diploid human fibroblasts." Science **277**(5327): 831-834.

Brugarolas, J., C. Chandrasekaran, et al. (1995). "Radiation-induced cell cycle arrest compromised by p21 deficiency." Nature **377**(6549): 552-557.

Buas, M. F., S. Kabak, et al. (2010). "The Notch effector Hey1 associates with myogenic target genes to repress myogenesis." J Biol Chem **285**(2): 1249-1258.

Buchholz, T. A., M. M. Weil, et al. (2004). "A Ser49Cys variant in the ataxia telangiectasia, mutated, gene that is more common in patients with breast carcinoma compared with population controls." Cancer **100**(7): 1345-1351.

Buchman, V. L., P. M. Chumakov, et al. (1988). "A variation in the structure of the protein-coding region of the human p53 gene." Gene **70**(2): 245-252.

Canman, C. E., D. S. Lim, et al. (1998). "Activation of the ATM kinase by ionizing radiation and phosphorylation of p53." Science **281**(5383): 1677-1679.

Chang, B. L., S. L. Zheng, et al. (2004). "A polymorphism in the CDKN1B gene is associated with increased risk of hereditary prostate cancer." Cancer Res **64**(6): 1997-1999.

Chedid, M., P. Michieli, et al. (1994). "A single nucleotide substitution at codon 31 (Ser/Arg) defines a polymorphism in a highly conserved region of the p53-inducible gene WAF1/CIP1." Oncogene **9**(10): 3021-3024.

Chen, G. K. and D. C. Thomas (2010). "Using biological knowledge to discover higher order interactions in genetic association studies." Genet Epidemiol **34**(8): 863-878.

Chen, J., D. Li, et al. (2008). "Polymorphisms of p16, p27, p73, and MDM2 Modulate Response and Survival of Pancreatic Cancer Patients Treated with Preoperative Chemoradiation." Ann Surg Oncol **16**(2):431-9.

Chen, P. H., Y. P. Tsao, et al. (2008). "Nuclear receptor interaction protein, a coactivator of androgen receptors (AR), is regulated by AR and Sp1 to feed forward and activate its own gene expression through AR protein stability." Nucleic Acids Res **36**(1): 51-66.

Chen, W., J. D. Arroyo, et al. (2005). "Cancer-associated PP2A Aalpha subunits induce functional haploinsufficiency and tumorigenicity." Cancer Res **65**(18): 8183-8192.

Chen, W. C., H. C. Wu, et al. (2002). "p21 gene codon 31 polymorphism is associated with bladder cancer." Urol Oncol **7**(2): 63-66.

Chen, X., E. M. Sturgis, et al. (2008). "p73 G4C14-to-A4T14 polymorphism and risk of human papillomavirus-associated squamous cell carcinoma of the oropharynx in never smokers and never drinkers." Cancer **113**(12): 3307-3314.

Cheng, Q., L. Chen, et al. (2009). "ATM activates p53 by regulating MDM2 oligomerization and E3 processivity." EMBO J **28**(24): 3857-3867.

Cheng, R., B. L. Ford, et al. (1997). "Zebrafish (Danio rerio) p53 tumor suppressor gene: cDNA sequence and expression during embryogenesis." Mol Mar Biol Biotechnol **6**(2): 88-97.

Chew, S. L., L. Baginsky, et al. (2000). "An exonic splicing silencer in the testes-specific DNA ligase III beta exon." Nucleic Acids Res **28**(2): 402-410.
Choi, J. E., H. G. Kang, et al. (2006). "No association between p73 G4C14-to-A4T14 polymorphism and the risk of lung cancer in a Korean population." Biochem Genet **44**(11-12): 543-550.
Chompret, A., L. Brugieres, et al. (2000). "P53 germline mutations in childhood cancers and cancer risk for carrier individuals." Br J Cancer **82**(12): 1932-1937.
Chu, I. M., L. Hengst, et al. (2008). "The Cdk inhibitor p27 in human cancer: prognostic potential and relevance to anticancer therapy." Nat Rev Cancer **8**(4): 253-267.
Committee_on_Bioethics (2001). "Ethical issues with genetic testing in pediatrics." Pediatrics **107**(6): 1451-1455.
Concannon, P. (2002). "ATM heterozygosity and cancer risk." Nat Genet **32**(1): 89-90.
Corbishley, T. P., M. J. Iqbal, et al. (1986). "Androgen receptor in human normal and malignant pancreatic tissue and cell lines." Cancer **57**(10): 1992-1995.
Cox, A., A. M. Dunning, et al. (2007). "A common coding variant in CASP8 is associated with breast cancer risk." Nat Genet **39**(3): 352-358.
Czernik, P. J., C. A. Peterson, et al. (1996). "Preferential binding of MyoD-E12 versus myogenin-E12 to the murine sarcoma virus enhancer in vitro." J Biol Chem **271**(15): 9141-9149.
Damo, L. A., P. W. Snyder, et al. (2005). "Tumorigenesis in p27/p53- and p18/p53-double null mice: functional collaboration between the pRb and p53 pathways." Mol Carcinog **42**(2): 109-120.
Danovi, D., E. Meulmeester, et al. (2004). "Amplification of Mdmx (or Mdm4) directly contributes to tumor formation by inhibiting p53 tumor suppressor activity." Mol Cell Biol **24**(13): 5835-5843.
Darbre, P. D. and R. J. King (1988). "Steroid hormone regulation of cultured breast cancer cells." Cancer Treat Res **40**: 307-341.
De Feo, E., R. Persiani, et al. (2009). "A case-control study on the effect of p53 and p73 gene polymorphisms on gastric cancer risk and progression." Mutat Res **675**(1-2): 60-65.
De Marzo, A. M., C. Bradshaw, et al. (1998). "CD44 and CD44v6 downregulation in clinical prostatic carcinoma: relation to Gleason grade and cytoarchitecture." Prostate **34**(3): 162-168.
Deng, C., P. Zhang, et al. (1995). "Mice lacking p21CIP1/WAF1 undergo normal development, but are defective in G1 checkpoint control." Cell **82**(4): 675-684.
Derry, W. B., A. P. Putzke, et al. (2001). "Caenorhabditis elegans p53: role in apoptosis, meiosis, and stress resistance." Science **294**(5542): 591-595.
Devlin, B. and N. Risch (1995). "A comparison of linkage disequilibrium measures for fine-scale mapping." Genomics **29**(2): 311-322.
Di Cristofano, A., B. Pesce, et al. (1998). "Pten is essential for embryonic development and tumour suppression." Nat Genet **19**(4): 348-355.
Dietel, M., M. A. Lewis, et al. (2005). "Hormone replacement therapy: pathobiological aspects of hormone-sensitive cancers in women relevant to epidemiological studies on HRT: a mini-review." Hum Reprod **20**(8): 2052-2060.

Dohoney, K. M., C. Guillerm, et al. (2004). "Phosphorylation of p53 at serine 37 is important for transcriptional activity and regulation in response to DNA damage." Oncogene **23**(1): 49-57.
Dombernowsky, S. L., M. Weischer, et al. (2008). "Risk of cancer by ATM missense mutations in the general population." J Clin Oncol **26**(18): 3057-3062.
Donehower, L. A., M. Harvey, et al. (1992). "Mice deficient for p53 are developmentally normal but susceptible to spontaneous tumours." Nature **356**(6366): 215-221.
Donehower, L. A. and G. Lozano (2009). "20 years studying p53 functions in genetically engineered mice." Nat Rev Cancer **9**(11): 831-841.
Dong, M., Y. Nio, et al. (2003). "p53 alteration is not an independent prognostic indicator, but affects the efficacy of adjuvant chemotherapy in human pancreatic cancer." J Surg Oncol **82**(2): 111-120.
Donis-Keller, H., P. Green, et al. (1987). "A genetic linkage map of the human genome." Cell **51**(2): 319-337.
Dou, J., M. Pan, et al. (2007). "Isolation and identification of cancer stem-like cells from murine melanoma cell lines." Cell Mol Immunol **4**(6): 467-472.
Driver, K. E., H. Song, et al. (2008). "Association of single-nucleotide polymorphisms in the cell cycle genes with breast cancer in the British population." Carcinogenesis **29**(2): 333-341.
Dumont, P., J. I. Leu, et al. (2003). "The codon 72 polymorphic variants of p53 have markedly different apoptotic potential." Nat Genet **33**(3): 357-365.
Durbin, R. M., G. R. Abecasis, et al. (2010). "A map of human genome variation from population-scale sequencing." Nature **467**(7319): 1061-1073.
Ehrhardt, H., S. Hacker, et al. (2008). "Cytotoxic drug-induced, p53-mediated upregulation of caspase-8 in tumor cells." Oncogene **27**(6): 783-793.
Enjuanes, A., Y. Benavente, et al. (2008). "Genetic variants in apoptosis and immunoregulation-related genes are associated with risk of chronic lymphocytic leukemia." Cancer Res **68**(24): 10178-10186.
Ernst, J., P. Kheradpour, et al. (2011). "Mapping and analysis of chromatin state dynamics in nine human cell types." Nature **473**(7345): 43-49.
Evans, D. G., P. Lunt, et al. (2010). "Childhood predictive genetic testing for Li-Fraumeni syndrome." Fam Cancer **9**(1): 65-69.
Fairbrother, W. G., R. F. Yeh, et al. (2002). "Predictive identification of exonic splicing enhancers in human genes." Science **297**(5583): 1007-1013.
Fairbrother, W. G., G. W. Yeo, et al. (2004). "RESCUE-ESE identifies candidate exonic splicing enhancers in vertebrate exons." Nucleic Acids Res **32**(Web Server issue): W187-190.
Fang, F., X. J. Yu, et al. (2010). "MDM2 309 T/G polymorphism is associated with colorectal cancer risk especially in Asians: a meta-analysis." Med Oncol [Epub ahead of print].
Fang, Y., B. Kong, et al. (2009). "MDM2 309 polymorphism is associated with missed abortion." Hum Reprod **24**(6): 1346-1349.
Feuk, L., A. R. Carson, et al. (2006). "Structural variation in the human genome." Nat Rev Genet **7**(2): 85-97.

Firouzabadi, R. D., N. Ghasemi, et al. (2009). "Association of p53 polymorphism with ICSI/IVF failure and recurrent pregnancy loss." Aust N Z J Obstet Gynaecol **49**(2): 216-219.

Firoz, E. F., M. Warycha, et al. (2009). "Association of MDM2 SNP309, age of onset, and gender in cutaneous melanoma." Clin Cancer Res **15**(7): 2573-2580.

Florenes, V. A., G. M. Maelandsmo, et al. (1994). "MDM2 gene amplification and transcript levels in human sarcomas: relationship to TP53 gene status." J Natl Cancer Inst **86**(17): 1297-1302.

Frank, A. K., J. I. Leu, et al. (2011). "The Codon 72 Polymorphism of p53 Regulates Interaction with NF-{kappa}B and Transactivation of Genes Involved in Immunity and Inflammation." Mol Cell Biol **31**(6): 1201-1213.

Frank, B., J. L. Bermejo, et al. (2005). "Re: Association of a common variant of the CASP8 gene with reduced risk of breast cancer." J Natl Cancer Inst **97**(13): 1012; author reply 1012-1013.

Frazer, K. A., D. G. Ballinger, et al. (2007). "A second generation human haplotype map of over 3.1 million SNPs." Nature **449**(7164): 851-861.

Frazer, K. A., S. S. Murray, et al. (2009). "Human genetic variation and its contribution to complex traits." Nat Rev Genet **10**(4): 241-251.

Freeman, D. J., A. G. Li, et al. (2003). "PTEN tumor suppressor regulates p53 protein levels and activity through phosphatase-dependent and -independent mechanisms." Cancer Cell **3**(2): 117-130.

Garcea, G., C. P. Neal, et al. (2005). "Molecular prognostic markers in pancreatic cancer: a systematic review." Eur J Cancer **41**(15): 2213-2236.

Garraway, L. A., H. R. Widlund, et al. (2005). "Integrative genomic analyses identify MITF as a lineage survival oncogene amplified in malignant melanoma." Nature **436**(7047): 117-122.

Giardiello, F. M., J. D. Brensinger, et al. (2000). "Very high risk of cancer in familial Peutz-Jeghers syndrome." Gastroenterology **119**(6): 1447-1453.

Godar, S., T. A. Ince, et al. (2008). "Growth-inhibitory and tumor- suppressive functions of p53 depend on its repression of CD44 expression." Cell **134**(1): 62-73.

Gonzalez, K. D., K. A. Noltner, et al. (2009). "Beyond Li Fraumeni Syndrome: clinical characteristics of families with p53 germline mutations." J Clin Oncol **27**(8): 1250-1256.

Grochola, L. F., T. H. Muller, et al. (2009). "MDM2 SNP309 Associates With Accelerated Pancreatic Adenocarcinoma Formation." Pancreas **39**(1):76-80.

Grochola, L. F., A. Vazquez, et al. (2009). "Recent Natural Selection Identifies a Genetic Variant in a Regulatory Subunit of Protein Phosphatase 2A that Associates with Altered Cancer Risk and Survival." Clin Cancer Res **15**(19):6301-8.

Grochola, L. F., J. Zeron-Medina, et al. (2010). "Single-nucleotide polymorphisms in the p53 signaling pathway." Cold Spring Harb Perspect Biol **2**(5): a001032.

Gryshchenko, I., S. Hofbauer, et al. (2008). "MDM2 SNP309 is associated with poor outcome in B-cell chronic lymphocytic leukemia." J Clin Oncol **26**(14): 2252-2257.

Gudas, J. M., H. Nguyen, et al. (1995). "Differential expression of multiple MDM2 messenger RNAs and proteins in normal and tumorigenic breast epithelial cells." Clin Cancer Res **1**(1): 71-80.

Hachiya, T., Y. Kuriaki, et al. (1999). "WAF1 genotype and endometrial cancer susceptibility." Gynecol Oncol **72**(2): 187-192.
Hall, J. M., J. F. Couse, et al. (2001). "The multifaceted mechanisms of estradiol and estrogen receptor signaling." J Biol Chem **276**(40): 36869-36872.
Harima, Y., S. Sawada, et al. (2001). "Polymorphism of the WAF1 gene is related to susceptibility to cervical cancer in Japanese women." Int J Mol Med **7**(3): 261-264.
Harris, N., E. Brill, et al. (1986). "Molecular basis for heterogeneity of the human p53 protein." Mol Cell Biol **6**(12): 4650-4656.
Hartmann, L. C., D. J. Schaid, et al. (1999). "Efficacy of bilateral prophylactic mastectomy in women with a family history of breast cancer." N Engl J Med **340**(2): 77-84.
Hengartner, M. O. (2000). "The biochemistry of apoptosis." Nature **407**(6805): 770-776.
Hidalgo, M. (2010). "Pancreatic cancer." N Engl J Med **362**(17): 1605-1617.
Hill, W. G. and B. S. Weir (1994). "Maximum-likelihood estimation of gene location by linkage disequilibrium." Am J Hum Genet **54**(4): 705-714.
Hingorani, S. R., L. Wang, et al. (2005). "Trp53R172H and KrasG12D cooperate to promote chromosomal instability and widely metastatic pancreatic ductal adenocarcinoma in mice." Cancer Cell **7**(5): 469-483.
Hirata, H., Y. Hinoda, et al. (2007). "MDM2 SNP309 Polymorphism as Risk Factor for Susceptibility and Poor Prognosis in Renal Cell Carcinoma." Clin Cancer Res **13**(14): 4123-4129.
Hisada, M., J. E. Garber, et al. (1998). "Multiple primary cancers in families with Li-Fraumeni syndrome." J Natl Cancer Inst **90**(8): 606-611.
Holst, F., P. R. Stahl, et al. (2007). "Estrogen receptor alpha (ESR1) gene amplification is frequent in breast cancer." Nat Genet **39**(5): 655-660.
Hong, Y., X. Miao, et al. (2005). "The role of P53 and MDM2 polymorphisms in the risk of esophageal squamous cell carcinoma." Cancer Res **65**(20): 9582-9587.
Houlston, R. S., E. Webb, et al. (2008). "Meta-analysis of genome-wide association data identifies four new susceptibility loci for colorectal cancer." Nat Genet **40**(12): 1426-1435.
Hu, W. (2009). "The role of p53 gene family in reproduction." Cold Spring Harb Perspect Biol **1**(6): a001073.
Hu, W., Z. Feng, et al. (2008). "p53: a new player in reproduction." Cell Cycle **7**(7): 848-852.
Hu, W., Z. Feng, et al. (2007). "A Single Nucleotide Polymorphism in the MDM2 Gene Disrupts the Oscillation of p53 and MDM2 Levels in Cells." Cancer Res **67**(6): 2757-2765.
Hu, W., Z. Feng, et al. (2007). "p53 regulates maternal reproduction through LIF." Nature **450**(7170): 721-724.
Hu, Y. F., B. Luscher, et al. (1990). "Transcription factor AP-4 contains multiple dimerization domains that regulate dimer specificity." Genes Dev **4**(10): 1741-1752.
Hu, Z., G. Jin, et al. (2007). "MDM2 promoter polymorphism SNP309 contributes to tumor susceptibility: evidence from 21 case-control studies." Cancer Epidemiol Biomarkers Prev **16**(12): 2717-2723.

Hu, Z., X. Miao, et al. (2005). "Dinucleotide polymorphism of p73 gene is associated with a reduced risk of lung cancer in a Chinese population." Int J Cancer **114**(3): 455-460.
Huang, S. F., I. H. Chen, et al. (2008). "Combined effects of MDM2 SNP 309 and p53 mutation on oral squamous cell carcinomas associated with areca quid chewing." Oral Oncol.
Huang, S. P., W. J. Wu, et al. (2004). "p53 Codon 72 and p21 codon 31 polymorphisms in prostate cancer." Cancer Epidemiol Biomarkers Prev **13**(12): 2217-2224.
Huarte, M., M. Guttman, et al. (2010). "A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response." Cell **142**(3): 409-419.
Hurt, E. M., B. T. Kawasaki, et al. (2008). "CD44+ CD24(-) prostate cells are early cancer progenitor/stem cells that provide a model for patients with poor prognosis." Br J Cancer **98**(4): 756-765.
Hwang, S. J., G. Lozano, et al. (2003). "Germline p53 mutations in a cohort with childhood sarcoma: sex differences in cancer risk." Am J Hum Genet **72**(4): 975-983.
Ikediobi, O. N., H. Davies, et al. (2006). "Mutation analysis of 24 known cancer genes in the NCI-60 cell line set." Mol Cancer Ther **5**(11): 2606-2612.
Iles, M. M. (2008). "What can genome-wide association studies tell us about the genetics of common disease?" PLoS Genet **4**(2): e33.
International_HapMap_Consortium (2003). "The International HapMap Project." Nature **426**(6968): 789-796.
International_HapMap_Consortium (2005). "A haplotype map of the human genome." Nature **437**(7063): 1299-1320.
Issa, J. P., B. A. Zehnbauer, et al. (1996). "The estrogen receptor CpG island is methylated in most hematopoietic neoplasms." Cancer Res **56**(5): 973-977.
Jin, S., S. Martinek, et al. (2000). "Identification and characterization of a p53 homologue in Drosophila melanogaster." Proc Natl Acad Sci U S A **97**(13): 7301-7306.
Jin, Z., L. Wallace, et al. (2010). "PP2A:B56{epsilon}, a substrate of caspase-3, regulates p53-dependent and p53-independent apoptosis during development." J Biol Chem **285**(45): 34493-34502.
Johnson, G. G., P. D. Sherrington, et al. (2009). "A novel type of p53 pathway dysfunction in chronic lymphocytic leukemia resulting from two interacting single nucleotide polymorphisms within the p21 gene." Cancer Res **69**(12): 5210-5217.
Johnstone, R. W., A. A. Ruefli, et al. (2002). "Apoptosis: a link between cancer genetics and chemotherapy." Cell **108**(2): 153-164.
Jones, S. N., A. E. Roe, et al. (1995). "Rescue of embryonic lethality in Mdm2-deficient mice by absence of p53." Nature **378**(6553): 206-208.
Jordan, V. C. (2006). "Tamoxifen (ICI46,474) as a targeted therapy to treat and prevent breast cancer." Br J Pharmacol **147 Suppl 1**: S269-276.
Joshi, A. M., S. Budhathoki, et al. (2011). "TP53 R72P and MDM2 SNP309 polymorphisms and colorectal cancer risk: the Fukuoka Colorectal Cancer Study." Jpn J Clin Oncol **41**(2): 232-238.
Jun, H. J., S. H. Park, et al. (2007). "Combined effects of p73 and MDM2 polymorphisms on the risk of lung cancer." Mol Carcinog **46**(2): 100-105.

Jung, P., A. Menssen, et al. (2008). "AP4 encodes a c-MYC-inducible repressor of p21." Proc Natl Acad Sci U S A **105**(39): 15046-15051.

Junttila, M. R., P. Puustinen, et al. (2007). "CIP2A inhibits PP2A in human malignancies." Cell **130**(1): 51-62.

Kaghad, M., H. Bonnet, et al. (1997). "Monoallelically expressed gene related to p53 at 1p36, a region frequently deleted in neuroblastoma and other human cancers." Cell **90**(4): 809-819.

Kanehisa, M. (1997) "A database for post-genome analysis." Trends Genet 13(9): 375-376.

Kang, H. J., Z. Feng, et al. (2009). "Single-nucleotide polymorphisms in the p53 pathway regulate fertility in humans." Proc Natl Acad Sci U S A **106**(24): 9761-9766.

Kastan, M. B. and D. S. Lim (2000). "The many substrates and functions of ATM." Nat Rev Mol Cell Biol **1**(3): 179-186.

Kay, C., R. S. Jeyendran, et al. (2006). "p53 tumour suppressor gene polymorphism is associated with recurrent implantation failure." Reprod Biomed Online **13**(4): 492-496.

Kelley, J. L. and W. J. Swanson (2008). "Positive selection in the human genome: from genome scans to biological significance." Annu Rev Genomics Hum Genet **9**: 143-160.

Kelley, M. L., P. Winge, et al. (2001). "Expression of homologues for p53 and p73 in the softshell clam (Mya arenaria), a naturally-occurring model for human cancer." Oncogene **20**(6): 748-758.

Keshava, C., B. L. Frye, et al. (2002). "Waf-1 (p21) and p53 polymorphisms in breast cancer." Cancer Epidemiol Biomarkers Prev **11**(1): 127-130.

Khatri, R. G., K. Navaratne, et al. (2008). "The role of a single nucleotide polymorphism of MDM2 in glioblastoma multiforme." J Neurosurg **109**(5): 842-848.

Khosravi, R., R. Maya, et al. (1999). "Rapid ATM-dependent phosphorylation of MDM2 precedes p53 accumulation in response to DNA damage." Proc Natl Acad Sci U S A **96**(26): 14973-14977.

Kim, J. W., H. Monila, et al. (2007). "Upstream stimulatory factors regulate the C/EBP alpha gene during differentiation of 3T3-L1 preadipocytes." Biochem Biophys Res Commun **354**(2): 517-521.

Kimura, S. H. and H. Nojima (2002). "Cyclin G1 associates with MDM2 and regulates accumulation and degradation of p53 protein." Genes Cells **7**(8): 869-880.

Kinyamu, H. K. and T. K. Archer (2003). "Estrogen receptor-dependent proteasomal degradation of the glucocorticoid receptor is coupled to an increase in mdm2 protein expression." Mol Cell Biol **23**(16): 5867-5881.

Klenova, E., A. C. Scott, et al. (2004). "YB-1 and CTCF differentially regulate the 5-HTT polymorphic intron 2 enhancer which predisposes to a variety of neurological disorders." J Neurosci **24**(26): 5966-5973.

Knappskog, S., M. Bjornslett, et al. (2011). "The MDM2 promoter SNP285C/309G haplotype diminishes Sp1 transcription factor binding and reduces risk for breast and ovarian cancer in Caucasians." Cancer Cell **19**(2): 273-282.

Koch, C. M., R. M. Andrews, et al. (2007). "The landscape of histone modifications across 1% of the human genome in five human cell lines." Genome Res **17**(6): 691-707.

Konduri, S., M. A. Schwarz, et al. (2007). "Androgen receptor blockade in experimental combination therapy of pancreatic cancer." J Surg Res **142**(2): 378-386.
Konduri, S. D., R. Medisetty, et al. (2010). "Mechanisms of estrogen receptor antagonism toward p53 and its implications in breast cancer therapeutic response and stem cell regulation." Proc Natl Acad Sci U S A **107**(34): 15081-15086.
Kortlever, R. M., P. J. Higgins, et al. (2006). "Plasminogen activator inhibitor-1 is a critical downstream target of p53 in the induction of replicative senescence." Nat Cell Biol **8**(8): 877-884.
Kos, M., G. Reid, et al. (2001). "Minireview: genomic organization of the human ERalpha gene promoter region." Mol Endocrinol **15**(12): 2057-2063.
Kraemer, K. H., M. M. Lee, et al. (1994). "The role of sunlight and DNA repair in melanoma and nonmelanoma skin cancer. The xeroderma pigmentosum paradigm." Arch Dermatol **130**(8): 1018-1021.
Ku, C. S., E. Y. Loy, et al. (2010). "The discovery of human genetic variations and their use as disease markers: past, present and future." J Hum Genet **55**(7): 403-415.
Kulkarni, D. A., A. Vazquez, et al. (2009). "A polymorphic variant in human MDM4 associates with accelerated age of onset of estrogen receptor negative breast cancer." Carcinogenesis **30**(11): 1910-1915.
Lahat, G., D. Tuvin, et al. (2008). "New perspectives for staging and prognosis in soft tissue sarcoma." Ann Surg Oncol **15**(10): 2739-2748.
Lai, J. C., Y. W. Cheng, et al. (2005). "Gender difference in estrogen receptor alpha promoter hypermethylation and its prognostic value in non-small cell lung cancer." Int J Cancer **117**(6): 974-980.
Lalloo, F., J. Varley, et al. (2003). "Prediction of pathogenic mutations in patients with early-onset breast cancer by family history." Lancet **361**(9363): 1101-1102.
Landa, I., C. Montero-Conde, et al. (2010). "Allelic variant at -79 (C>T) in CDKN1B (p27Kip1) confers an increased risk of thyroid cancer and alters mRNA levels." Endocr Relat Cancer **17**(2): 317-328.
Landa, I., S. Ruiz-Llorente, et al. (2009). "The variant rs1867277 in FOXE1 gene confers thyroid cancer susceptibility through the recruitment of USF1/USF2 transcription factors." PLoS Genet **5**(9): e1000637.
Lander, E. S., L. M. Linton, et al. (2001). "Initial sequencing and analysis of the human genome." Nature **409**(6822): 860-921.
Lane, D. and A. Levine (2010). "p53 Research: the past thirty years and the next thirty years." Cold Spring Harb Perspect Biol **2**(12): a000893.
Lane, D. P. and L. V. Crawford (1979). "T antigen is bound to a host protein in SV40-transformed cells." Nature **278**(5701): 261-263.
Lang, G. A., T. Iwakuma, et al. (2004). "Gain of function of a p53 hot spot mutation in a mouse model of Li-Fraumeni syndrome." Cell **119**(6): 861-872.
Leal, J. F., J. Fominaya, et al. (2008). "Cellular senescence bypass screen identifies new putative tumor suppressor genes." Oncogene **27**(14): 1961-1970.
Lee, C. J., J. Dosch, et al. (2008). "Pancreatic cancer stem cells." J Clin Oncol **26**(17): 2806-2812.
Lee, K. E., Y. S. Hong, et al. "p73 G4C14 to A4T14 polymorphism is associated with colorectal cancer risk and survival." World J Gastroenterol **16**(35): 4448-4454.

Leu, J. D., C. Y. Wang, et al. (2011). "Involvement of p53 R72P polymorphism in the association of MDM2-SNP309 with breast cancer." Oncol Rep **25**(6): 1755-1763.
Levine, A. J., W. Hu, et al. (2006). "The P53 pathway: what questions remain to be explored?" Cell Death Differ **13**(6): 1027-1036.
Li, A. G., L. G. Piluso, et al. (2006). "Mechanistic insights into maintenance of high p53 acetylation by PTEN." Mol Cell **23**(4): 575-587.
Li, F. P. (1995). "Identification and management of inherited cancer susceptibility." Environ Health Perspect **103 Suppl 8**: 297-300.
Li, F. P., J. F. Fraumeni, Jr., et al. (1988). "A cancer family syndrome in twenty-four kindreds." Cancer Res **48**(18): 5358-5362.
Li, F. P., L. C. Strong, et al. (1990). "Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms." Science **250**(4985): 1233-1238.
Li, G., Z. Liu, et al. (2005). "Genetic polymorphisms of p21 are associated with risk of squamous cell carcinoma of the head and neck." Carcinogenesis **26**(9): 1596-1602.
Li, G., E. M. Sturgis, et al. (2004). "Association of a p73 exon 2 G4C14-to-A4T14 polymorphism with risk of squamous cell carcinoma of the head and neck." Carcinogenesis **25**(10): 1911-1916.
Li, G., L. E. Wang, et al. (2004). "p73 G4C14-to-A4T14 polymorphism and risk of lung cancer." Cancer Res **64**(19): 6863-6866.
Li, H. H., X. Cai, et al. (2007). "A specific PP2A regulatory subunit, B56gamma, mediates DNA damage-induced dephosphorylation of p53 at Thr55." Embo J **26**(2): 402-411.
Liaw, D., D. J. Marsh, et al. (1997). "Germline mutations of the PTEN gene in Cowden disease, an inherited breast and thyroid cancer syndrome." Nat Genet **16**(1): 64-67.
Liedtke, C., N. Groger, et al. (2003). "The human caspase-8 promoter sustains basal activity through SP1 and ETS-like transcription factors and can be up-regulated by a p53-dependent mechanism." J Biol Chem **278**(30): 27593-27604.
Limacher, J. M., T. Frebourg, et al. (2001). "Two metachronous tumors in the radiotherapy fields of a patient with Li-Fraumeni syndrome." Int J Cancer **96**(4): 238-242.
Linares, L. K., A. Hengstermann, et al. (2003). "HdmX stimulates Hdm2-mediated ubiquitination and degradation of p53." Proc Natl Acad Sci U S A **100**(21): 12009-12014.
Lind, H., S. Zienolddiny, et al. (2006). "Association of a functional polymorphism in the promoter of the MDM2 gene with risk of nonsmall cell lung cancer." Int J Cancer **119**(3): 718-721.
Link, D. C., L. G. Schuettpelz, et al. (2011). "Identification of a novel TP53 cancer susceptibility mutation through whole-genome sequencing of a patient with therapy-related AML." Jama **305**(15): 1568-1576.
Linzer, D. I. and A. J. Levine (1979). "Characterization of a 54K dalton cellular SV40 tumor antigen present in SV40-transformed cells and uninfected embryonal carcinoma cells." Cell **17**(1): 43-52.

Liu, F., B. Li, et al. (2011) "P21 codon 31 polymorphism associated with cancer among white people: evidence from a meta-analysis involving 78 074 subjects." Mutagenesis 26(4):513-21.
Liu, G., T. J. McDonnell, et al. (2000). "High metastatic potential in mice inheriting a targeted p53 missense mutation." Proc Natl Acad Sci U S A **97**(8): 4174-4179.
Liu, J. and G. Jiang (2006). "CD44 and hematologic malignancies." Cell Mol Immunol **3**(5): 359-365.
Liu, W., M. M. Ip, et al. (2009). "Disruption of estrogen receptor alpha-p53 interaction in breast tumors: a novel mechanism underlying the anti-tumor effect of radiation therapy." Breast Cancer Res Treat **115**(1): 43-50.
Liu, W., S. D. Konduri, et al. (2006). "Estrogen receptor-alpha binds p53 tumor suppressor protein directly and represses its function." J Biol Chem **281**(15): 9837-9840.
Liu, Z., G. Li, et al. (2010). "Genetic variations in TERT-CLPTM1L genes and risk of squamous cell carcinoma of the head and neck." Carcinogenesis **31**(11): 1977-1981.
Logsdon, C. D., D. M. Simeone, et al. (2003). "Molecular profiling of pancreatic adenocarcinoma and chronic pancreatitis identifies multiple genes differentially regulated in pancreatic cancer." Cancer Res **63**(10): 2649-2657.
Lowe, S. W. and A. W. Lin (2000). "Apoptosis in cancer." Carcinogenesis **21**(3): 485-495.
Lozano, G. (2010). "Mouse models of p53 functions." Cold Spring Harb Perspect Biol **2**(4): a001115.
Lozano, G. and G. P. Zambetti (2005). "What have animal models taught us about the p53 pathway?" J Pathol **205**(2): 206-220.
Lu, M., R. Mira-y-Lopez, et al. (2005). "Expression of estrogen receptor alpha, retinoic acid receptor alpha and cellular retinoic acid binding protein II genes is coordinately regulated in human breast cancer cells." Oncogene **24**(27): 4362-4369.
Lu, S., G. Jenster, et al. (2000). "Androgen induction of cyclin-dependent kinase inhibitor p21 gene: role of androgen receptor and transcription factor Sp1 complex." Mol Endocrinol **14**(5): 753-760.
Lu, X. (2010). "Tied up in loops: positive and negative autoregulation of p53." Cold Spring Harb Perspect Biol **2**(5): a000984.
Lynch, H. T., R. M. Fusaro, et al. (2008). "Pancreatic cancer and the FAMMM syndrome." Fam Cancer **7**(1): 103-112.
Ma, H., G. Jin, et al. (2006). "Variant genotypes of CDKN1A and CDKN1B are associated with an increased risk of breast cancer in Chinese women." Int J Cancer **119**(9): 2173-2178.
MacCallum, C. and E. Hill (2006). "Being positive about selection." PLoS Biol **4**(3): e87.
MacKenzie, A. and J. Quinn (1999). "A serotonin transporter gene intron 2 polymorphic region, correlated with affective disorders, has allele-dependent differential enhancer-like properties in the mouse embryo." Proc Natl Acad Sci U S A **96**(26): 15251-15255.

Macleod, K. F., N. Sherry, et al. (1995). "p53-dependent and independent expression of p21 during cell growth, differentiation, and DNA damage." Genes Dev **9**(8): 935-944.
MacPherson, G., C. S. Healey, et al. (2004). "Association of a common variant of the CASP8 gene with reduced risk of breast cancer." J Natl Cancer Inst **96**(24): 1866-1869.
Malkin, D., F. P. Li, et al. (1990). "Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms." Science **250**(4985): 1233-1238.
Mancini, F., G. Di Conza, et al. (2009). "MDM4 (MDMX) and its Transcript Variants." Curr Genomics **10**(1): 42-50.
Manfredi, J. J. (2010). "The Mdm2-p53 relationship evolves: Mdm2 swings both ways as an oncogene and a tumor suppressor." Genes Dev **24**(15): 1580-1589.
Marcel, V., E. I. Palmero, et al. (2009). "TP53PIN3 and MDM2 SNP309 polymorphisms as genetic modifiers in the Li-Fraumeni syndrome: impact on age at first diagnosis." J Med Genet **46**(11):766-72.
Marchetti, A., F. Buttitta, et al. (1995). "mdm2 gene alterations and mdm2 protein expression in breast carcinomas." J Pathol **175**(1): 31-38.
Marin, M. C., C. A. Jost, et al. (2000). "A common polymorphism acts as an intragenic modifier of mutant p53 behaviour." Nat Genet **25**(1): 47-54.
Marine, J. C., S. Francoz, et al. (2006). "Keeping p53 in check: essential and synergistic functions of Mdm2 and Mdm4." Cell Death Differ **13**(6): 927-934.
Marsh, D. J., P. L. Dahia, et al. (1997). "Germline mutations in PTEN are present in Bannayan-Zonana syndrome." Nat Genet **16**(4): 333-334.
Martens, E., I. Stevens, et al. (2004). "Genomic organisation, chromosomal localisation tissue distribution and developmental regulation of the PR61/B' regulatory subunits of protein phosphatase 2A in mice." J Mol Biol **336**(4): 971-986.
Matushansky, I., E. Charytonowicz, et al. (2009). "MFH classification: differentiating undifferentiated pleomorphic sarcoma in the 21st Century." Expert Rev Anticancer Ther **9**(8): 1135-1144.
Maya, R., M. Balass, et al. (2001). "ATM-dependent phosphorylation of Mdm2 on serine 395: role in p53 activation by DNA damage." Genes Dev **15**(9): 1067-1077.
Mayo, L. D. and D. B. Donner (2001). "A phosphatidylinositol 3-kinase/Akt pathway promotes translocation of Mdm2 from the cytoplasm to the nucleus." Proc Natl Acad Sci U S A **98**(20): 11598-11603.
Mehta, M. S., A. Vazquez, et al. (2011). "Polymorphic variants in TSC1 and TSC2 and their association with breast cancer phenotypes." Breast Cancer Res Treat **125**(3): 861-868.
Mendrysa, S. M., M. K. McElwee, et al. (2003). "mdm2 Is critical for inhibition of p53 during lymphopoiesis and the response to ionizing irradiation." Mol Cell Biol **23**(2): 462-472.
Mendrysa, S. M., K. A. O'Leary, et al. (2006). "Tumor suppression and normal aging in mice with constitutively high p53 activity." Genes Dev **20**(1): 16-21.
Menin, C., M. C. Scaini, et al. (2006). "Association between MDM2-SNP309 and age at colorectal cancer diagnosis according to p53 mutation status." J Natl Cancer Inst **98**(4): 285-288.

Metivier, R., G. Penot, et al. (2003). "Estrogen receptor-alpha directs ordered, cyclical, and combinatorial recruitment of cofactors on a natural target promoter." Cell **115**(6): 751-763.
Migliaccio, A., M. Di Domenico, et al. (1996). "Tyrosine kinase/p21ras/MAP-kinase pathway activation by estradiol-receptor complex in MCF-7 cells." EMBO J **15**(6): 1292-1300.
Migliorini, D., E. Lazzerini Denchi, et al. (2002). "Mdm4 (Mdmx) regulates p53-induced growth arrest and neuronal cell death during early embryonic mouse development." Mol Cell Biol **22**(15): 5527-5538.
Miyoshi, H., M. Nakau, et al. (2002). "Gastrointestinal hamartomatous polyposis in Lkb1 heterozygous knockout mice." Cancer Res **62**(8): 2261-2266.
Molife, R., P. Lorigan, et al. (2001). "Gender and survival in malignant tumours." Cancer Treat Rev **27**(4): 201-209.
Moll, U. M., S. Wolff, et al. (2005). "Transcription-independent pro-apoptotic functions of p53." Curr Opin Cell Biol **17**(6): 631-636.
Momand, J., D. Jung, et al. (1998). "The MDM2 gene amplification database." Nucleic Acids Res **26**(15): 3453-3459.
Momand, J., H. H. Wu, et al. (2000). "MDM2--master regulator of the p53 tumor suppressor protein." Gene **242**(1-2): 15-29.
Montes de Oca Luna, R., D. S. Wagner, et al. (1995). "Rescue of early embryonic lethality in mdm2-deficient mice by deletion of p53." Nature **378**(6553): 203-206.
Morrison, D. K. (2009). "The 14-3-3 proteins: integrators of diverse signaling cues that impact cell fate and cancer development." Trends Cell Biol **19**(1): 16-23.
Moule, M. G., C. H. Collins, et al. (2004). "Role for PP2A in ARF signaling to p53." Proc Natl Acad Sci U S A **101**(39): 14063-14066.
Mumby, M. (2007). "PP2A: unveiling a reluctant tumor suppressor." Cell **130**(1): 21-24.
Murphy, K. M., K. A. Brune, et al. (2002). "Evaluation of candidate genes MAP2K4, MADH4, ACVR1B, and BRCA2 in familial pancreatic cancer: deleterious BRCA2 mutations in 17%." Cancer Res **62**(13): 3789-3793.
Murray-Zmijewski, F., D. P. Lane, et al. (2006). "p53/p63/p73 isoforms: an orchestra of isoforms to harmonise cell differentiation and response to stress." Cell Death Differ **13**(6): 962-972.
Myers, R. M., J. Stamatoyannopoulos, et al. (2011). "A user's guide to the encyclopedia of DNA elements (ENCODE)." PLoS Biol **9**(4): e1001046.
Nakamura, T., Y. Furukawa, et al. (2004). "Genome-wide cDNA microarray analysis of gene expression profiles in pancreatic cancers using populations of tumor cells and normal ductal epithelial cells selected for purity by laser microdissection." Oncogene **23**(13): 2385-2400.
Nakashima, M., S. Kondo, et al. (2008). "Impact of MDM2 single nucleotide polymorphism on tumor onset in head and neck squamous cell carcinoma." Acta Otolaryngol **128**(7): 808-813.
Naor, D., S. B. Wallach-Dayan, et al. (2008). "Involvement of CD44, a molecule with a thousand faces, in cancer dissemination." Semin Cancer Biol **18**(4): 260-267.
Nayak, M. S., J. M. Yang, et al. (2007). "Effect of a single nucleotide polymorphism in the murine double minute 2 promoter (SNP309) on the sensitivity to topoisomerase II-targeting drugs." Cancer Res **67**(12): 5831-5839.

Neale, B. M. (2010). "Introduction to linkage disequilibrium, the HapMap, and imputation." Cold Spring Harb Protoc **2010**(3): pdb top74.
Neupane, D. and M. Korc (2008). "14-3-3sigma Modulates pancreatic cancer cell survival and invasiveness." Clin Cancer Res **14**(23): 7614-7623.
Neve, R. M., K. Chin, et al. (2006). "A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes." Cancer Cell **10**(6): 515-527.
Neviani, P., R. Santhanam, et al. (2005). "The tumor suppressor PP2A is functionally inactivated in blast crisis CML through the inhibitory activity of the BCR/ABL-regulated SET protein." Cancer Cell **8**(5): 355-368.
Niwa, Y., K. Hirose, et al. (2005). "Association of p73 G4C14-to-A4T14 polymorphism at exon 2 and p53 Arg72Pro polymorphism with the risk of endometrial cancer in Japanese subjects." Cancer Lett **219**(2): 183-190.
Noureddine, M. A., D. Menendez, et al. (2009). "Probing the functional impact of sequence variation on p53-DNA interactions using a novel microsphere assay for protein-DNA binding with human cell extracts." PLoS Genet **5**(5): e1000462.
O'Connor, P. M., J. Jackman, et al. (1997). "Characterization of the p53 tumor suppressor pathway in cell lines of the National Cancer Institute anticancer drug screen and correlations with the growth-inhibitory potency of 123 anticancer agents." Cancer Res **57**(19): 4285-4300.
Ogawara, Y., S. Kishishita, et al. (2002). "Akt enhances Mdm2-mediated ubiquitination and degradation of p53." J Biol Chem **277**(24): 21843-21850.
Okamoto, K., H. Li, et al. (2002). "Cyclin G recruits PP2A to dephosphorylate Mdm2." Mol Cell **9**(4): 761-771.
Okumura, N., S. Saji, et al. (2002). "Distinct promoter usage of mdm2 gene in human breast cancer." Oncol Rep **9**(3): 557-563.
Olive, K. P., D. A. Tuveson, et al. (2004). "Mutant p53 gain of function in two mouse models of Li-Fraumeni syndrome." Cell **119**(6): 847-860.
Ollmann, M., L. M. Young, et al. (2000). "Drosophila p53 is a structural and functional homolog of the tumor suppressor p53." Cell **101**(1): 91-101.
Orlova, K. A. and P. B. Crino (2010). "The tuberous sclerosis complex." Ann N Y Acad Sci **1184**: 87-105.
Ottaviano, Y. L., J. P. Issa, et al. (1994). "Methylation of the estrogen receptor gene CpG island marks loss of estrogen receptor expression in human breast cancer cells." Cancer Res **54**(10): 2552-2555.
Palanca Suela, S., E. Esteban Cardenosa, et al. (2009). "CASP8 D302H polymorphism delays the age of onset of breast cancer in BRCA1 and BRCA2 carriers." Breast Cancer Res Treat **119**(1):87-93.
Pankow, S. and C. Bamberger (2007). "The p53 tumor suppressor-like protein nvp63 mediates selective germ cell death in the sea anemone Nematostella vectensis." PLoS One **2**(9): e782.
Parant, J., A. Chavez-Reyes, et al. (2001). "Rescue of embryonic lethality in Mdm4-null mice by loss of Trp53 suggests a nonoverlapping pathway with MDM2 to regulate p53." Nat Genet **29**(1): 92-95.
Park, S. H., J. E. Choi, et al. (2006). "MDM2 309T>G polymorphism and risk of lung cancer in a Korean population." Lung Cancer **54**(1): 19-24.

Pellegata, N. S., F. Sessa, et al. (1994). "K-ras and p53 gene mutations in pancreatic cancer: ductal and nonductal tumors progress through different genetic lesions." Cancer Res **54**(6): 1556-1560.

Pennacchio, L. A. and E. M. Rubin (2001). "Genomic strategies to identify mammalian regulatory sequences." Nat Rev Genet **2**(2): 100-109.

Peter, M. E. (2000). "The TRAIL DISCussion: It is FADD and caspase-8!" Cell Death Differ **7**(9): 759-760.

Pfeifer, D., G. Arbman, et al. (2005). "Polymorphism of the p73 gene in relation to colorectal cancer risk and survival." Carcinogenesis **26**(1): 103-107.

Phang, B. H. and K. Sabapathy (2007). "The codon 72 polymorphism-specific effects of human p53 are absent in mouse cells: implications on generation of mouse models." Oncogene **26**(21): 2964-2974.

Phelps, M., M. Darley, et al. (2003). "p53-independent activation of the hdm2-P2 promoter through multiple transcription factor response elements results in elevated hdm2 expression in estrogen receptor alpha-positive breast cancer cells." Cancer Res **63**(10): 2616-2623.

Philipp-Staheli, J., K. H. Kim, et al. (2004). "Distinct roles for p53, p27Kip1, and p21Cip1 during tumor development." Oncogene **23**(4): 905-913.

Phillips, A., A. Teunisse, et al. (2010). "HDMX-L is expressed from a functional p53-responsive promoter in the first intron of the HDMX gene and participates in an autoregulatory feedback loop to control p53 activity." J Biol Chem **285**(38): 29111-29127.

Phillips, C. L., R. Gerbing, et al. (2010). "MDM2 polymorphism increases susceptibility to childhood acute myeloid leukemia: a report from the Children's Oncology Group." Pediatr Blood Cancer **55**(2): 248-253.

Pietras, R. J. and D. C. Marquez-Garban (2007). "Membrane-associated estrogen receptor signaling pathways in human cancers." Clin Cancer Res **13**(16): 4672-4676.

Pim, D. and L. Banks (2004). "p53 polymorphic variants at codon 72 exert different effects on cell cycle progression." Int J Cancer **108**(2): 196-199.

Pittman, A. M., P. Broderick, et al. (2008). "CASP8 variants D302H and -652 6N ins/del do not influence the risk of colorectal cancer in the United Kingdom population." Br J Cancer **98**(8): 1434-1436.

Pittman, A. M., A. J. Myers, et al. (2005). "Linkage disequilibrium fine mapping and haplotype association analysis of the tau gene in progressive supranuclear palsy and corticobasal degeneration." J Med Genet **42**(11): 837-846.

Ponta, H., L. Sherman, et al. (2003). "CD44: from adhesion molecules to signalling regulators." Nat Rev Mol Cell Biol **4**(1): 33-45.

Post, S. M., A. Quintas-Cardama, et al. (2010). "A high-frequency regulatory polymorphism in the p53 pathway accelerates tumor development." Cancer Cell **18**(3): 220-230.

Rada-Iglesias, A., R. Bajpai, et al. (2011). "A unique chromatin signature uncovers early developmental enhancers in humans." Nature **470**(7333): 279-283.

Rallapalli, R., G. Strachan, et al. (1999). "A novel MDMX transcript expressed in a variety of transformed cell lines encodes a truncated protein with potent p53 repressive activity." J Biol Chem **274**(12): 8299-8308.

Ramos, E. A., A. A. Camargo, et al. (2010). "Simultaneous CXCL12 and ESR1 CpG island hypermethylation correlates with poor prognosis in sporadic breast cancer." BMC Cancer **10**: 23.

Raponi, M., J. Kralovicova, et al. (2011). "Prediction of single-nucleotide substitutions that result in exon skipping: identification of a splicing silencer in BRCA1 exon 6." Hum Mutat **32**(4): 436-444.

Redston, M. S., C. Caldas, et al. (1994). "p53 mutations in pancreatic carcinoma and evidence of common involvement of homocopolymer tracts in DNA microdeletions." Cancer Res **54**(11): 3025-3033.

Reed, J. C. (1999). "Dysregulation of apoptosis in cancer." J Clin Oncol **17**(9): 2941-2953.

Reinbold, M., J. L. Luo, et al. (2008). "Common tumour p53 mutations in immortalized cells from Hupki mice heterozygous at codon 72." Oncogene **27**(19): 2788-2794.

Riancho, J. A., C. Garcia-Ibarbia, et al. "Common variations in estrogen-related genes are associated with severe large-joint osteoarthritis: a multicenter genetic and functional study." Osteoarthritis Cartilage **18**(7):927-33.

Riley, T., E. Sontag, et al. (2008). "Transcriptional control of human p53-regulated genes." Nat Rev Mol Cell Biol **9**(5): 402-412.

Rippe, R. A., S. I. Lorenzen, et al. (1989). "Regulatory elements in the 5'-flanking region and the first intron contribute to transcriptional control of the mouse alpha 1 type I collagen gene." Mol Cell Biol **9**(5): 2224-2227.

Rodrigues, F. C., R. S. Kawasaki-Oyama, et al. (2003). "Analysis of CDKN1A polymorphisms: markers of cancer susceptibility?" Cancer Genet Cytogenet **142**(2): 92-98.

Roh, J., M. Kim, et al. (2001). "Polymorphisms in codon 31 of p21 and cervical cancer susceptibility in Korean women." Cancer Lett **165**(1): 59-62.

Roh, J. W., J. W. Kim, et al. (2004). "p53 and p21 genetic polymorphisms and susceptibility to endometrial cancer." Gynecol Oncol **93**(2): 499-505.

Rosenbloom, K. R., T. R. Dreszer, et al. (2010). "ENCODE whole-genome data in the UCSC Genome Browser." Nucleic Acids Res **38**(Database issue): D620-625.

Rotter, V., D. Schwartz, et al. (1993). "Mice with reduced levels of p53 protein exhibit the testicular giant-cell degenerative syndrome." Proc Natl Acad Sci U S A **90**(19): 9075-9079.

Rowland, B. D. and D. S. Peeper (2006). "KLF4, p21 and context-dependent opposing forces in cancer." Nat Rev Cancer **6**(1): 11-23.

Ruijs, M. W., M. K. Schmidt, et al. (2007). "The single-nucleotide polymorphism 309 in the MDM2 gene contributes to the Li-Fraumeni syndrome and related phenotypes." Eur J Hum Genet **15**(1): 110-114.

Runkel, L., P. E. Shaw, et al. (1991). "Multiple basal promoter elements determine the level of human c-fos transcription." Mol Cell Biol **11**(3): 1270-1280.

Ryan, B. M., R. McManus, et al. (2001). "A common p73 polymorphism is associated with a reduced incidence of oesophageal carcinoma." Br J Cancer **85**(10): 1499-1503.

Sabeti, P. C., D. E. Reich, et al. (2002). "Detecting recent positive selection in the human genome from haplotype structure." Nature **419**(6909): 832-837.

Sablina, A. A., W. Chen, et al. (2007). "The tumor suppressor PP2A Abeta regulates the RalA GTPase." Cell **129**(5): 969-982.

Sachidanandam, R., D. Weissman, et al. (2001). "A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms." Nature **409**(6822): 928-933.

Saito, S., A. A. Goodarzi, et al. (2002). "ATM mediates phosphorylation at multiple p53 sites, including Ser(46), in response to ionizing radiation." J Biol Chem **277**(15): 12491-12494.

Sakamuro, D., P. Sabbatini, et al. (1997). "The polyproline region of p53 is required to activate apoptosis but not growth arrest." Oncogene **15**(8): 887-898.

Salvioli, S., M. Bonafe, et al. (2005). "p53 codon 72 alleles influence the response to anticancer drugs in cells from aged people by regulating the cell cycle inhibitor p21WAF1." Cell Cycle **4**(9): 1264-1271.

Sanchez-Carbayo, M., N. D. Socci, et al. (2007). "A polymorphism in HDM2 (SNP309) associates with early onset in superficial tumors, TP53 mutations, and poor outcome in invasive bladder cancer." Clin Cancer Res **13**(11): 3215-3220.

Savitsky, K., A. Bar-Shira, et al. (1995). "A single ataxia telangiectasia gene with a product similar to PI-3 kinase." Science **268**(5218): 1749-1753.

Sayeed, A., S. D. Konduri, et al. (2007). "Estrogen receptor alpha inhibits p53-mediated transcriptional repression: implications for the regulation of apoptosis." Cancer Res **67**(16): 7746-7755.

Scarpa, A., P. Capelli, et al. (1993). "Pancreatic adenocarcinomas frequently show p53 gene mutations." Am J Pathol **142**(5): 1534-1543.

Schmitt, C. A., C. T. Rosenthal, et al. (2000). "Genetic analysis of chemoresistance in primary murine lymphomas." Nat Med **6**(9): 1029-1035.

Schmitt, C. A., R. R. Wallace-Brodeur, et al. (2000). "DNA damage responses and chemosensitivity in the E mu-myc mouse lymphoma model." Cold Spring Harb Symp Quant Biol **65**: 499-510.

Schneider, K. and J. Garber (1993-1999). "Li-Fraumeni Syndrome." In: Pagon, R. A., T. D: Bird, C. R. Dolan, K. Stephens, editors. SourceGeneReviews [Internet]. Seattle (WA): University of Washington, Seattle [updated 2010 Feb 09]

Schuit, S. C., H. H. Oei, et al. (2004). "Estrogen receptor alpha gene polymorphisms and risk of myocardial infarction." Jama **291**(24): 2969-2977.

Schuit, S. C., J. B. van Meurs, et al. (2004). "Height in pre- and postmenopausal women is influenced by estrogen receptor alpha gene polymorphisms." J Clin Endocrinol Metab **89**(1): 303-309.

J. L. Sebaugh (2010). “Guidelines for accurate EC50/IC50 estimation.” Pharmaceutical statistics **10**(2): 128-134.

Shankar, S., D. Nall, et al. (2011). "Resveratrol inhibits pancreatic cancer stem cell characteristics in human and kras transgenic mice by inhibiting pluripotency maintaining factors and epithelial-mesenchymal transition." PLoS One **6**(1): e16530.

Sheikh, M. S., Z. M. Shao, et al. (1993). "The p53-binding protein MDM2 gene is differentially expressed in human breast carcinoma." Cancer Res **53**(14): 3226-3228.

Shen, L., Y. Kondo, et al. (2007). "Drug sensitivity prediction by CpG island methylation profile in the NCI-60 cancer cell line panel." Cancer Res **67**(23): 11335-11343.
Sherry, S. T., M. Ward, et al. (1999). "dbSNP-database for single nucleotide polymorphisms and other classes of minor genetic variation." Genome Res **9**(8): 677-679.
Sherry, S. T., M. H. Ward, et al. (2001). "dbSNP: the NCBI database of genetic variation." Nucleic Acids Res **29**(1): 308-311.
Shi, H., S. J. Tan, et al. (2009). "Winter temperature and UV are tightly linked to genetic changes in the p53 tumor suppressor pathway in Eastern Asia." Am J Hum Genet **84**(4): 534-541.
Shiloh, Y. (2003). "ATM and related protein kinases: safeguarding genome integrity." Nat Rev Cancer **3**(3): 155-168.
Shoemaker, R. H. (2006). "The NCI60 human tumour cell line anticancer drug screen." Nat Rev Cancer **6**(10): 813-823.
Shtivelman, E. and J. M. Bishop (1991). "Expression of CD44 is repressed in neuroblastoma cells." Mol Cell Biol **11**(11): 5446-5453.
Sigurdson, A. J., P. Bhatti, et al. (2007). "Polymorphisms in apoptosis- and proliferation-related genes, ionizing radiation exposure, and risk of breast cancer among U.S. Radiologic Technologists." Cancer Epidemiol Biomarkers Prev **16**(10): 2000-2007.
Simoncini, T., A. Hafezi-Moghadam, et al. (2000). "Interaction of oestrogen receptor with the regulatory subunit of phosphatidylinositol-3-OH kinase." Nature **407**(6803): 538-541.
Sjalander, A., R. Birgander, et al. (1996). "Association between the p21 codon 31 A1 (arg) allele and lung cancer." Hum Hered **46**(4): 221-225.
Slatkin, M. (2008). "Linkage disequilibrium--understanding the evolutionary past and mapping the medical future." Nat Rev Genet **9**(6): 477-485.
Smirnov, D. A., M. Morley, et al. (2009). "Genetic analysis of radiation-induced changes in human gene expression." Nature **459**(7246): 587-591.
Soussi, T., C. Caron de Fromentel, et al. (1987). "Cloning and characterization of a cDNA from Xenopus laevis coding for a protein homologous to human and murine p53." Oncogene **1**(1): 71-78.
Stadler, Z. K., P. Thom, et al. (2010). "Genome-wide association studies of cancer." J Clin Oncol **28**(27): 4255-4267.
Stavridi, E. S., N. H. Chehab, et al. (2001). "Substitutions that compromise the ionizing radiation-induced association of p53 with 14-3-3 proteins also compromise the ability of p53 to induce cell cycle arrest." Cancer Res **61**(19): 7030-7033.
Stratton, M. R., P. J. Campbell, et al. (2009). "The cancer genome." Nature **458**(7239): 719-724.
Stredrick, D. L., M. Garcia-Closas, et al. (2006). "The ATM missense mutation p.Ser49Cys (c.146C>G) and the risk of breast cancer." Hum Mutat **27**(6): 538-544.
Su, L., G. Liu, et al. (2003). "No association between the p21 codon 31 serine-arginine polymorphism and lung cancer risk." Cancer Epidemiol Biomarkers Prev **12**(2): 174-175.

Su, L., Y. Sai, et al. (2003). "P53 (codon 72) and P21 (codon 31) polymorphisms alter in vivo mRNA expression of p21." Lung Cancer **40**(3): 259-266.
Sullivan, A., N. Syed, et al. (2004). "Polymorphism in wild-type p53 modulates response to chemotherapy in vitro and in vivo." Oncogene **23**(19): 3328-3337.
Sultana, A., C. Tudur Smith, et al. (2007). "Systematic review, including meta-analyses, on the management of locally advanced pancreatic cancer using radiation/combined modality therapy." Br J Cancer **96**(8): 1183-1190.
Sun, Y., A. Hildesheim, et al. (1995). "No point mutation but a codon 31ser-->arg polymorphism of the WAF-1/CIP-1/p21 tumor suppressor gene in nasopharyngeal carcinoma (NPC): the polymorphism distinguishes Caucasians from Chinese." Cancer Epidemiol Biomarkers Prev **4**(3): 261-267.
Suzuki, H. I., K. Yamagata, et al. (2009). "Modulation of microRNA processing by p53." Nature **460**(7254): 529-533.
Swinney, R. M., S. C. Hsu, et al. (2005). "MDM2 promoter variation and age of diagnosis of acute lymphoblastic leukemia." Leukemia **19**(11): 1996-1998.
Szeszel-Fedorowicz, W., I. Talukdar, et al. (2006). "An exonic splicing silencer is involved in the regulated splicing of glucose 6-phosphate dehydrogenase mRNA." J Biol Chem **281**(45): 34146-34158.
Taubert, H., A. Meye, et al. (1998). "Soft tissue sarcomas and p53 mutations." Mol Med **4**(6): 365-372.
Taubert, H., P. Wurl, et al. (1998). "[Frequency, distribution and prognostic relevance of p53 mutations in soft tissue sarcomas]." Verh Dtsch Ges Pathol **82**: 340-344.
Terzian, T., Y. Wang, et al. (2007). "Haploinsufficiency of Mdm2 and Mdm4 in Tumorigenesis and Development." Mol Cell Biol **27**(15):5479-85.
The_International_HapMap_Consortium (2005). "A haplotype map of the human genome." Nature **437**(7063): 1299-1320.
Thomas, G., K. B. Jacobs, et al. (2009). "A multistage genome-wide association study in breast cancer identifies two new risk alleles at 1p11.2 and 14q24.1 (RAD51L1)." Nat Genet **41**(5): 579-584.
Thomas, G., K. B. Jacobs, et al. (2008). "Multiple loci identified in a genome-wide association study of prostate cancer." Nat Genet **40**(3): 310-315.
Thomas, M., A. Kalita, et al. (1999). "Two polymorphic variants of wild-type p53 differ biochemically and biologically." Mol Cell Biol **19**(2): 1092-1100.
Thompson, E. W., S. Paik, et al. (1992). "Association of increased basement membrane invasiveness with absence of estrogen receptor and expression of vimentin in human breast cancer cell lines." J Cell Physiol **150**(3): 534-544.
Thull, D. L. and V. G. Vogel (2004). "Recognition and management of hereditary breast cancer syndromes." Oncologist **9**(1): 13-24.
Tilley, W. D., C. M. Wilson, et al. (1990). "Androgen receptor gene expression in human prostate carcinoma cell lines." Cancer Res **50**(17): 5382-5386.
Toffoli, G., P. Biason, et al. (2009). "Effect of TP53 Arg72Pro and MDM2 SNP309 polymorphisms on the risk of high-grade osteosarcoma development and survival." Clin Cancer Res **15**(10): 3550-3556.
Toledo, F. and G. M. Wahl (2006). "Regulating the p53 pathway: in vitro hypotheses, in vivo veritas." Nat Rev Cancer **6**(12): 909-923.

Toledo, F. and G. M. Wahl (2007). "MDM2 and MDM4: p53 regulators as targets in anticancer therapy." Int J Biochem Cell Biol **39**(7-8):1476-82.

Tomso, D. J., A. Inga, et al. (2005). "Functionally distinct polymorphic sequences in the human genome that are targets for p53 transactivation." Proc Natl Acad Sci U S A **102**(18): 6431-6436.

Tsujimoto, K., T. Ono, et al. (2005). "Regulation of the expression of caspase-9 by the transcription factor activator protein-4 in glucocorticoid-induced apoptosis." J Biol Chem **280**(30): 27638-27644.

Tzivion, G., V. S. Gupta, et al. (2006). "14-3-3 proteins as potential oncogenes." Semin Cancer Biol **16**(3): 203-213.

Ueda, M., M. Yamamoto, et al. (2009). "Murine double-minute 2 homolog single nucleotide polymorphism 309 and the risk of gynecologic cancer." Hum Cell **22**(2): 49-54.

van Meurs, J. B., S. C. Schuit, et al. (2003). "Association of 5' estrogen receptor alpha gene polymorphisms with bone mineral density, vertebral bone area and fracture risk." Hum Mol Genet **12**(14): 1745-1754.

Varley, J. M. (2003). "Germline TP53 mutations and Li-Fraumeni syndrome." Hum Mutat **21**(3): 313-320.

Varley, J. M., D. G. Evans, et al. (1997). "Li-Fraumeni syndrome--a molecular and clinical review." Br J Cancer **76**(1): 1-14.

Vazquez, A., E. E. Bond, et al. (2008). "The genetics of the p53 pathway, apoptosis and cancer therapy." Nat Rev Drug Discov **7**(12): 979-987.

Vazquez, A., L. F. Grochola, et al. (2010). "Chemosensitivity profiles identify polymorphisms in the p53 network genes 14-3-3tau and CD44 that affect sarcoma incidence and survival." Cancer Res **70**(1): 172-180.

Vazquez, A., D. Kulkarni, et al. (2011). "A genetic variant in a PP2A regulatory subunit encoded by the PPP2R2B gene associates with altered breast cancer risk and recurrence." Int J Cancer **128**(10): 2335-2343.

Venkatachalam, S., S. D. Tyner, et al. (2001). "Is p53 haploinsufficient for tumor suppression? Implications for the p53+/- mouse model in carcinogenicity testing." Toxicol Pathol **29 Suppl**: 147-154.

Venter, J. C., M. D. Adams, et al. (2001). "The sequence of the human genome." Science **291**(5507): 1304-1351.

Vikhanskaya, F., M. M. Siddique, et al. (2005). "Evaluation of the combined effect of p53 codon 72 polymorphism and hotspot mutations in response to anticancer drugs." Clin Cancer Res **11**(12): 4348-4356.

Villani, A., U. Tabori, et al. (2011). "Biochemical and imaging surveillance in germline TP53 mutation carriers with Li-Fraumeni syndrome: a prospective observational study." Lancet Oncol **12**(6): 559-567.

Voight, B. F., S. Kudaravalli, et al. (2006). "A map of recent positive selection in the human genome." PLoS Biol **4**(3): e72.

Vousden, K. H. and C. Prives (2009). "Blinded by the Light: The Growing Complexity of p53." Cell **137**(3): 413-431.

Wade, M., Y. V. Wang, et al. (2010). "The p53 orchestra: Mdm2 and Mdmx set the tone." Trends Cell Biol **20**(5): 299-309.

Walker, K. K. and A. J. Levine (1996). "Identification of a novel p53 functional domain that is necessary for efficient growth suppression." Proc Natl Acad Sci U S A **93**(26): 15335-15340.
Wallace-Brodeur, R. R. and S. W. Lowe (1999). "Clinical implications of p53 mutations." Cell Mol Life Sci **55**(1): 64-75.
Wan, Y., W. Wu, et al. (2011). "MDM2 SNP309, gene-gene interaction, and tumor susceptibility: an updated meta-analysis." BMC Cancer **11**(1): 208.
Wang, B., K. Liu, et al. (2004). "A role for 14-3-3 tau in E2F1 stabilization and DNA damage-induced apoptosis." J Biol Chem **279**(52): 54140-54152.
Wang, D. Z., M. R. Valdez, et al. (2001). "The Mef2c gene is a direct transcriptional target of myogenic bHLH and MEF2 proteins during skeletal muscle development." Development **128**(22): 4623-4633.
Wang, K., W. D. Li, et al. (2011). "A genome-wide association study on obesity and obesity-related traits." PLoS One **6**(4): e18939.
Wang, L., R. Gao, et al. (2011). "Combined analysis of the association between p73 G4C14-to-A4T14 polymorphisms and cancer risk." Mol Biol Rep [Epub ahead of print].
Wang, R., R. G. Crystal, et al. (2009). "Identification of an exonic splicing silencer in exon 6A of the human VEGF gene." BMC Mol Biol **10**: 103.
Wang, Z. and C. B. Burge (2008). "Splicing regulation: from a parts list of regulatory elements to an integrated splicing code." RNA **14**(5): 802-813.
Wang, Z., M. E. Rolish, et al. (2004). "Systematic identification and analysis of exonic splicing silencers." Cell **119**(6): 831-845.
Wasielewski, M., J. H. Nagel, et al. (2006). "MDM2 SNP309 accelerates familial breast carcinogenesis independently of estrogen signaling." Breast Cancer Res Treat **104**(2):153-7.
Waterman, M. J., E. S. Stavridi, et al. (1998). "ATM-dependent activation of p53 involves dephosphorylation and association with 14-3-3 proteins." Nat Genet **19**(2): 175-178.
Weber, G. F., R. T. Bronson, et al. (2002). "Absence of the CD44 gene prevents sarcoma metastasis." Cancer Res **62**(8): 2281-2286.
Wedren, S., L. Lovmar, et al. (2008). "Estrogen receptor alpha gene polymorphism and endometrial cancer risk--a case-control study." BMC Cancer **8**: 322.
Wei, M. C., W. X. Zong, et al. (2001). "Proapoptotic BAX and BAK: a requisite gateway to mitochondrial dysfunction and death." Science **292**(5517): 727-730.
Weickert, C. S., A. L. Miranda-Angulo, et al. (2008). "Variants in the estrogen receptor alpha gene and its mRNA contribute to risk for schizophrenia." Hum Mol Genet **17**(15): 2293-2309.
Welboren, W. J., M. A. van Driel, et al. (2009). "ChIP-Seq of ERalpha and RNA polymerase II defines genes differentially responding to ligands." EMBO J **28**(10): 1418-1428.
Westermarck, J. and W. C. Hahn (2008). "Multiple pathways regulated by the tumor suppressor PP2A in transformation." Trends Mol Med **14**(4): 152-160.
Whibley, C., P. D. Pharoah, et al. (2009). "p53 polymorphisms: cancer implications." Nat Rev Cancer **9**(2): 95-107.

Widschwendter, M., K. D. Siegmund, et al. (2004). "Association of breast cancer DNA methylation profiles with hormone receptor status and response to tamoxifen." Cancer Res **64**(11): 3807-3813.

Wielenga, V. J., K. H. Heider, et al. (1993). "Expression of CD44 variant proteins in human colorectal cancer is related to tumor progression." Cancer Res **53**(20): 4754-4756.

Wright, M. H., A. M. Calcagno, et al. (2008). "Brca1 breast tumors contain distinct CD44+/CD24- and CD133+ cells with cancer stem cell characteristics." Breast Cancer Res **10**(1): R10.

Wu, C. C., R. Krahe, et al. (2011). "Joint effects of germ-line TP53 mutation, MDM2 SNP309, and gender on cancer risk in family studies of Li-Fraumeni syndrome." Hum Genet **129**(6): 663-673.

Wu, M. T., D. C. Wu, et al. (2003). "Association between p21 codon 31 polymorphism and esophageal cancer risk in a Taiwanese population." Cancer Lett **201**(2): 175-180.

Wynendaele, J., A. Bohnke, et al. (2010). "An illegitimate microRNA target site within the 3' UTR of MDM4 affects ovarian cancer progression and chemosensitivity." Cancer Res **70**(23): 9641-9649.

Xirodimas, D. P., M. K. Saville, et al. (2004). "Mdm2-mediated NEDD8 conjugation of p53 inhibits its transcriptional activity." Cell **118**(1): 83-97.

Xu, Y. (2003). "Regulation of p53 responses by post-translational modifications." Cell Death Differ **10**(4): 400-403.

Yao, Z., S. Duan, et al. (2007). "Death effector domain DEDa, a self-cleaved product of caspase-8/Mch5, translocates to the nucleus by binding to ERK1/2 and upregulates procaspase-8 expression via a p53-dependent mechanism." Embo J **26**(4): 1068-1080.

Yeo, T. P., R. H. Hruban, et al. (2002). "Pancreatic cancer." Curr Probl Cancer **26**(4): 176-275.

Yoon, Y. J., H. Y. Chang, et al. (2008). "MDM2 and p53 Polymorphisms are Associated with the Development of Hepatocellular Carcinoma in Patients with Chronic Hepatitis B Virus Infection." Carcinogenesis **29**(6):1192-6.

Yu, Q., B. P. Toole, et al. (1997). "Induction of apoptosis of metastatic mammary carcinoma cells in vivo by disruption of tumor cell surface CD44 function." J Exp Med **186**(12): 1985-1996.

Zhang, L., J. Yu, et al. (2000). "Role of BAX in the apoptotic response to anticancer agents." Science **290**(5493): 989-992.

Zhang, X., P. J. Nicholls, et al. (2010). "A functional alternative splicing mutation in human tryptophan hydroxylase-2." Mol Psychiatry [Epub ahead of print].

Zhou, B. B. and S. J. Elledge (2000). "The DNA damage response: putting checkpoints in perspective." Nature **408**(6811): 433-439.

Zhou, B. P., Y. Liao, et al. (2001). "HER-2/neu induces p53 ubiquitination via Akt-mediated MDM2 phosphorylation." Nat Cell Biol **3**(11): 973-982.

Zhou, M., L. Gu, et al. (2003). "PTEN reverses MDM2-mediated chemotherapy resistance by interacting with p53 in acute lymphoblastic leukemia cells." Cancer Res **63**(19): 6357-6362.

Zhu, F., M. E. Dolle, et al. (2010). "Mouse models for the p53 R72P polymorphism mimic human phenotypes." Cancer Res **70**(14): 5851-5859.
Zilfou, J. T. and S. W. Lowe (2009). "Tumor suppressive functions of p53." Cold Spring Harb Perspect Biol **1**(5): a001883.