Identification and characterization of TNFα responsive genes in human breast cancer cells

Dissertation
zur Erlangung des Doktorgrades
der Naturwissenschaften

vorgelegt beim Fachbereich Biologie und Informatik
der Johann Wolfgang Goethe – Universität
in Frankfurt am Main

von Francisca Guardiola Serrano
aus Palma de Mallorca, Spanien

Frankfurt am Main 2006
D30

vom Fachbereich Biologie und Mathematik der Johann Wolfgang Goethe – Universität als Dissertation angenommen

Dekan: Prof. Dr. Rüdiger Wittig

Gutachter: Prof. Dr. Anna Starzinski-Powitz

Prof. Dr. Harald von Melchner

Datum der Disputation:

Acknowledgments:

The present study was carried out at the Department of Hematology of the Johann Wolfgang Goethe University Hospital in Frankfurt. There is lot of people I would like to thank for a huge variety of reasons.

I am deeply grateful to Prof. Dr. Harald von Melchner for allowing me to do a Ph.D. thesis in his department and for kindly accepting and supporting my work.

I am also very grateful to Prof. Dr. Anna Starzinski-Powitz for taking the part of responsibility for this thesis and for being external referee.

I owe most sincere gratitude and respect to my group leader Dr. Joachim Altschmied. He spent a lot of his valuable time on my research problems and was always ready and patient to guide and advise me. I could not have imagined having a better advisor and mentor for my PhD, and without his common-sense, knowledge, perceptiveness, everlasting enthusiasm, contagious optimism and kindness I would never have finished.

I owe my special thanks to Mrs. Maria Doris Viel for her kindness and help with all the administrative work.

My warmest thanks go to the other members of the molecular hematology team who made me feel truly accepted and shared both happy and unhappy laboratory days with me during these years.

I am very grateful to my dear colleagues and friends Karsten Sturm, Silke De-Zolt, Armin Refelman, Annoesjka Goessen and specially Corinna Strolz for amusing coffee breaks, bringing fun in the laboratory life and offering help for all kind of problems.

I wish to extend my acknowledgements to my dear friends outside the laboratory: Carinna, Kerstin and Manuela for amusing times that we shared. And especially to Corinna Strolz and Hilal Gül for their valuable advices and endless encouragement. I am very grateful to the friends who support me from the distance (Spain): Marga Puigserver, Lucia Moreno and Marlen Morey for their nice visits, valuable advices as well as optimism.

I also would like to express my gratitude to the *GlaxoSmithKline Stiftung* for the travel grant that allowed me to attend the CSHL Symposium.

Finally, I would like to express my deepest gratitude to my parents and sister for their love, patient, trust, never-ending support and encouragement during this study and throughout my life. I would never complete my studies without them. I dedicate my thesis to my family members.

Translation of my gratitude for my family: Mi profunda gratitud es para mis padres y mi hermana por su amor, confianza, paciencia y apoyo durante el presente estudio y a través de la vida. Sin ellos esta tesis no habría sido posible. A ellos les dedico mi tesis.

Abbreviations:

%	Percent
µg	Microgram
µg/ml	Micromol per mililiter
µl	Microliter
µM	Micromol per liter
µm	Micrometer
ABCC3	ATP-binding cassette, sub-family C (CFTR/MRP), member 3
ABL	Abelson murine leukemia virus homology gene
ADP	Adenosine diphosphate
AMPO	Aminopeptidase O
APAF-1	Apoptotic protease activating factor
ARF	ADP-ribosylation factor
ARHGAP11A	Rho GTPase activating protein 11A
Ash1	(absent, small, or homeotic)-like (Drosophila)
ATP	Adenosine-5′ tryphosphate
Bad	BCL-2 antagonist of cell death
BAFF	B cell activating factor belonging to the TNF family
BAFFR	BAFF receptor
Bax	BCL-2 associated X protein
BCL9L	B-cell CLL/lymphoma 9-like
BCL-X_L	B-cell lymphoma X_L
BCR	B-cell receptor
BCR	Breakpoint cluster region
bGHpA	Bovine growth hormone polyadenylation sequence
BID	BH3 interacting domain death agonist
BIR	Baculovirus IAP repeat
bla	Beta lactamase
bp	Base pair
BRCA2	Breast cancer 2 tumor suppressor
BSA	Bovine serum albumin
C1QTNF6	C1q and tumor necrosis factor related protein 6
C20orf142	Chromosome 20 open reading frame 142
C9orf3	Chromosome 9 open reading frame 3
CASP-10/-3/-8	Caspase-10/-3/-8
cDNA	Copy DNA
Ci/ml	Curies per mililiter
Ci/mmol	Curies per milimol
CMV	Cytomegalovirus
Cre	*Cre* recombinase gene from bacteriophage P1
CTNND2	Catenin (cadherin-associated protein), delta 2 (neural plakophilin-related arm-repeat protein)

Cyt. c	Cytochrome c
DAPI	Diaminophenylindole
DAXX	Death associated protein 6
DEN	Dietylnitrosamine
DEPC	Diethylpyrocarbonate
DIABLO	Direct IAP-binding protein with low pI
DISC	Death-inducing signalling complex
DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethylsulfoxid
DNA	Desoxyribonucleic acid
dnFADD	Dominant negative FADD
dNTPs	Desoxiribonucleotides triphosphate
DSS	Dextran-sulphate sodium salt
DTT	Dithiothreitol
E. Coli	*Escherichia Coli*
e.g.	Exampli gratia
E_2	Estradiol
EBV	Epstein-Barr virus
ECL	Enhanced chemiluminescence
ECM	Extra Cellular Matrix
EDTA	ethylenediaminetetraacetic acid
EED	Embryonic Ectoderm development protein, mouse homolog of
EGF	Epidermal growth factor
EGFP	Enhanced green fluorescent protein
eIF-2α	Eukaryotic translation initiation factor 2-alpha
ESR	Estrogen receptor
ESR1	Estrogen receptor alpha
EST	Expressed Sequence Tags
FADD	Fas-associated death domain
FANTOM	Functional Annotation of the mouse
FCS	Foetal calf serum
FGD3	FYVE, RhoGEF and PH domain containing 3
FLIP	Flice-inhibitory protein
FLJ	Full-length Long Japan project
g	Units of gravity
g/l	Gram per liter
Gfp	Green fluorescent protein
GO	Gene ontology
GRHL3	Grainyhead, *Drosophila*, homolog of, 3
GTPase	Rho family guanosine triphosphatases
GTSTs	Gene Trap Sequence Tags
HBA2	α globin gene

HDAC-1	Histone deacetylase 1
HDACs	Histone deacetylases
HGF	Hepatocyte growth factor
HKII	Hexokinase 2
HNRPH1	Heterogeneous nuclear ribonucleoprotein H1 (H)
HPLC	High performance liquid chromatography
HSV	Herpes simplex virus
HTLV-1	T-cell leukemia virus
I.M.A.G.E.	Integrated Molecular Analysis of Genome and their Expression
IAPs	Inhibitor of apoptosis proteins
ie	Id est, that is
IKK	IκB kinase complex
IKK α/β/ γ	Inhibitor of NF-κB (IκB) kinase α/β/γ
IL-1/-6/-8	Interleukin -1/-6/-8
IL1-R	Interleukin 1 receptor
iPCR	Inverse PCR
IPTG	Isopropyl-beta-D-thiogalacto-pyranoside
IRES	Internal ribosomal entry site
IκBα/ IκBβ	Inhibitor of κBα/Inhibitor of κBβ
JNK	c-Jun N-terminal kinase
kbp	Kilobase pair
KCTD5	Potassium channel tetramerisation domain containing 5
kDa	Kilodalton
kg	Kilogram
KIF11/KNSL1	Kinesin family member 11/Kinesin-like 1
LB medium	Luria-Bertani-Medium
LDS	Lauryl dodecyl sulfate
LIPC	Lipase hepatic
LTβ	Lymphotoxin B
LTβR	Lymphotoxin B receptor
lx	*loxP* sites from bacteriophage P1
M	Mol per liter
MAPK	Mitogen-activated protein kinase
MAPKK	Mitogen-activated protein kinase kinase
MBD4	Methyl-CpG binding domain protein 4
MDR2	Multidrug resistance 2 gene
MEM	Minimum essential medium
mg/l	Miligrams per liter
mg/ml	Milligram per mililiter
MHC	Major histocompatibility complex
min	Minute
miRNAs	micro RNAs

ml	Mililiter
mM	Milimol per liter
MMLV	Moloney murine leukemia virus
MMPs	Matrix metaloproteinases
MOI	Multiplicity of infection
MOPS	3-morpholino-propane-sulfonic acid
mRNA	Messenger ribonucleic acid
NATs	Natural antisense transcripts
NB	Nuclear body
ncRNAs	Non-coding RNAs
neo	Neomycin phosphotransferase
NF-κB	Nuclear factor kappa B
ng	Nanogram
ng/ml	Nanograms per mililiter
NIK	NF-κB inducing kinase
NK-cell	Natural killer cell
nm	Nanometers
NNPP	Neural network promoter prediction
NTN4	Netrin 4
O.D.	Optical density
ORF	Open reading frame
PARD6B	Partitioning-defective protein 6, C.Elegans, homolog of, beta
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PGK	Phosphoglycerate-kinase
PITPNM2	Phosphatidylinositol transfer protein, membrane-associated 2
PML-1	Promyelocytic leukemia
pmol	Picomol
PMSF	Phenylmethylsulphonylfluoride
POK	POZ and Krüppel family of proteins
Polybrene	Hexadimethindibromide
PRKCBP-1	Protein kinase C binding protein 1
Pro-CASP-9/-3/-8	Pro-caspase -9/-3/-8
PRSS36	Protease, serine, 36
puro	puromycin-acetyl-transferase
REV-T	Reticuloentotheliosis retrovirus
RHD	Rel-homology domain
RING	really interesting new gene
RIP	Receptor interactive protein
RISC	RNA-induced silencing complex
RNA	Ribonucleic acid
RNF184	Ring finger protein 184
ROS	Reactive oxygen species

rpm	Revolutions per minute
RPMI medium	Roswell Park Memorial Institute-Medium
RZPD	German resource center for genome research
S100A10	S100 calcium binding protein A10 (annexin II ligand, calpactin I, light polypeptide (p11))
S100P	S100 calcium binding protein P
S100PBP	S100P binding protein
SDS	Sodium dodecyl sulfate
SDS-PAGE	SDS-polyacrylamid gel electrophoresis
sec	Second
SLC12A2	Solute carrier family 12 (sodium/potassium/chloride transporters), member 2
SLC1A2	Solute carrier family 1 (glial high affinity glutamate transporter), member 2
SLC7A2	Solute carrier family 7 (cationic amino acid transporter, y+ system), member 2
SMAC	Second mitochondria-derived activator of caspases
SODD	silencer of death domain protein
SOM/TFCP2L4	Sister of mammalian grainyhead/ transcription factor CP2 like 4
SP100	SP100 nuclear antigen
STAT-1	Signal transducer and activator of transcription 1
SUMO	Small ubiquitin like modifier
SV40	Simian virus 40
SV40epa	SV40 early polyadenylation sequence
TBS	Tris-buffered saline
TBST	Tris-buffered saline + Tween 20
TCR	T-cell receptor
TEMED	N,N,N',N'-Tetramethylethylendiamin
TEX14	Testis expressed sequence 14
TFF1	Trefoil factor 1 (breast cancer, estrogen-inducible sequence expressed in)
tkneo	Fusion between timidin kinase and neomycin-phosphotransferase
TNFR1	Tumor necrosis factor receptor type I
TNFα	Tumor necrosis factor α
TOP1	Topoisomerase (DNA) I
TRADD	TNF receptor associated death domain
TRAF2	TNF receptor-associated factor-2
Tris	Tris-(hydroximethyl)-aminomethan
TSS	Trancription start site
TWEAK	TNF-related weak inducer of apoptosis
U	Unit
U.V.	Ultraviolet light
USP40	Ubiquitin specific peptidase 40

UTR	Untranslated region
V	Voltage
v/v	Volume per volume
VEGA	Vertebrate and genome analysis
VSV-G	Veiscular Stomatitis virus G
w/v	Weight per volume
WDR10	WD-repeat protein 10
X-Gal	5-bromo-4-chloro-3-indolyl-b-d-galactopyranoside in N,N'-dimethyl-formamide
XIAP	X-chromosome-linked IAP
YARS	Tyrosyl-tRNA synthetase
ZFP67	Zinc finger protein 67, mouse homolog of
ZNF143	Zinc finger protein 143 (clone pHZ-1)
ZNHIT2	Zinc finger, HIT type 2
ZNRF1	Zinc and ring finger 1

Index

1. - Introduction 1

1.1. Hallmarks of cancer, the route to becoming a cancer cell 1

1.2. A link between inflammation and cancer 3

1.3. NF-κB a major player in cancer progression 5

1.3.1. The NF-κB transcription family 5

1.3.2. NF-κB as the link between cancer and inflammation 8

1.3.3. NF-κB activation in tumors of different origins 10

1.4. A genomewide screen for recovery of survival genes 12

1.4.1 TNFα as an inducer of NF-κB signaling 12

1.4.2. Combination of gene trap with Cre/*loxP* induced recombination allows detection of transiently TNFα–induced genes 16

1.5. Aim of the project 19

2. - Materials 20

2.1. Machines and technical devices 20

2.2. General laboratory supplies 21

2.3. Kits 21

2.4. Chemicals 21

2.5. Enzymes 22

2.5.1. Restriction enzymes 22

2.5.2. Modifying enzymes 22

2.5.3. Enzyme inhibitors 22

2.6. Pre-mixed solutions 22

2.7. Reagents 23

2.8. Cell culture medium 23

2.9. Cytokines 23

2.10. Antibiotics 23

2.11. Molecular weight standards: 23

2.12. Desoxyoligonucleotides 23

2.13. Nucleotides 23

2.14. Antibodies 24
2.15. Vectors 24
2.15.1. Commercial vectors 24
2.15.2 Provided vectors 24
2.15.3. Constructed vectors 24
2.16. Software 25
2.16.1. On line bioinformatics resources 25

3. - Methods 26
3.1. Purification, manipulation and detection of DNA 26
3.1.1. DNA storage 26
3.1.2. Plasmid DNA isolation 26
3.1.3. DNA isolation from eukaryotic cell lines 26
3.1.4. Measurement of DNA concentration 27
3.1.5. Enzymatic manipulation of DNA 27
3.1.5.1. Endonucleolytic digestion 27
3.1.5.2. Dephosphorylation 27
3.1.5.3. Ligation 27
3.1.5.4. Filling in of double stranded DNA with 5'-overhangs 27
3.1.6. Electrophoresis of DNA in agarose gels 27
3.1.7. DNA purification 28
3.1.8. Polymerase chain reaction (PCR) 28
3.1.9. Inverse PCR 29
3.1.10. Radioactive labeling of DNA fragments 30
3.2. Purification, manipulation and detection of RNA 30
3.2.2. RNA isolation from eukaryotic cell lines 30
3.2.3. Measurement of RNA concentration 31
3.2.4. RNA concentration 31
3.2.5. mRNA isolation 31
3.2.6. DNA digestion 31
3.2.7. First strand cDNA synthesis 32
3.2.8. Northern blot 32
3.2.8.1. RNA electrophoresis 32

3.2.8.1.1. Gel preparation ... 32
3.2.8.1.2. Sample preparation ... 32
3.2.8.2. Transfer to nitrocellulose membranes ... 33
3.2.8.3. Staining of RNA on nitrocellulose membranes ... 33
3.2.8.4. Hybridization ... 34
3.2.8.4.1. Hybridization probe ... 34
3.2.8.4.2. Hybridization to the membrane ... 34
3.2.8.5. Detection ... 35
3.2.8.6. Stripping ... 35
3.3. Preparation, storage and detection of proteins ... 35
3.3.2. Preparation of cell lysates/extracts ... 35
3.3.3. Measurement of protein concentration ... 36
3.3.4. Western blot ... 36
3.3.4.1. Sodium dodecyl sulfate polyacrylamid gel electrophoresis (SDS-PAGE) . 36
3.3.4.1.1. Gel preparation ... 36
3.3.4.1.2. Electrophoresis ... 37
3.3.4.2. Electroblotting ... 37
3.3.4.3. Staining ... 38
3.3.4.4. Immunological detection ... 38
3.4. Working with bacteria ... 40
3.4.1. Bacterial strains ... 40
3.4.2. Growth and storage ... 41
3.4.3. Competent bacteria ... 41
3.4.4. Transformation with DNA ... 42
3.5. Work with mammalian cells ... 43
3.5.1. Cell lines ... 43
3.5.2. Routine culturing ... 43
3.5.3. Cell counting ... 44
3.5.4. Freezing and thawing of cells ... 45
3.5.5. Transfection of mammalian cells ... 45
3.5.5.1. Transfection with Fu Gene 6 ... 45
3.5.5.2. Transfection with TransPass D1 ... 46
3.5.6. Clone production ... 46

3.5.7. Cell proliferation assay 46
3.5.8. Immunofluorescence 47
3.5.9. Growth in soft agar 47
3.6. Working with retrovirus 48
3.6.1. Gene transduction with retroviral vectors 48
3.6.1.1. Production of replication deficient, VSV-G pseudotyped retroviruses 48
3.6.1.2. Concentration of VSV-G pseudotyped MMLV-retroviruses 49
3.6.1.3. Infection of human cell lines with VSV-G pseudotyped retroviruses 49
3.6.2. Generation of the switch reporter cell line 49
3.6.3. Generation of the U3Cre integration library 50
3.6.3.1. Gene trap titration 50
3.6.3.2. Creation of the gene trap library and selection of cell clones with integrations in TNFα inducible genes 51

4. - Results 52
4.1. Identification of TNFα induced genes in MCF-7 cells 52
4.1.1. Design of a cell line reporting TNFα induced gene trap insertions 52
4.1.2. Stable over-expression of a dominant-negative FADD protects MCF-7 cells from TNFα induced cell death 54
4.1.3. Derivation of a MCF-7 reporter cell line suitable for trapping TNFα inducible genes 56
4.1.4. Isolation of cell clones with U3Cre integrations in TNFα inducible loci 59
4.1.5. Recovery of genomic gene trap sequence tags by inverse PCR 60
4.2. *In silico* analysis of GTSTs 62
4.2.1. Identification of integration sites 62
4.2.2. U3Cre insertions are mainly in 5'-introns 63
4.2.3. U3Cre traps putative antisense (non-coding?) transcripts 64
4.2.4. Recovered genes belong to different functional gene classes 72
4.2.5. Recovery of cancer-related genes 73
4.3. Validation of the TNFα inducible genes 77
4.3.1. Regulation by TNFα 77
4.3.2. Selection of candidate genes for functional characterization 80
4.3.3. Over-expression studies 81

4.3.4. Subcellular localization 83
4.3.5. Over-expression of FLJ14451 protein inhibits colony formation in soft agar 85

5. - Discussion 86

5.1. Recovery of TNFα induced genes by combined gene trap mutagenesis and site specific recombination 86
5.2. Mechanisms of entrapment 90
5.3. Nature of the recovered transcripts 95
5.4. Functional validation of the trapped genes 99
5.5. Future perspectives 100

6. - References 101

7. - Summary 115

8. – German Summary 117

9. - Annex 122

9.1. Plasmids maps 122
9.2. Primers 125
9.2.1. For probe amplification 125
9.2.2. For ORF amplification 127
9.2.3. For recombination test 127
9.2.4. For inverse PCR 127
9.3. Sequences obtained by inverse PCR 128
9.4. Bioinformatics analysis 134

1. - Introduction

1.1. Hallmarks of cancer, the route to becoming a cancer cell

Coordination and balance between cell proliferation and programmed cell death (apoptosis) is crucial for normal development and homeostasis in metazoan organisms. During early development only few cells abandon the cell cycle, whereas in an adult organism most of the cells are quiescent. In these cells, a GAP called G1 phase is incorporated between nuclear division (M phase) and DNA synthesis (S phase), another GAP called G2 occurred between S and M phases. During G1 the cell makes further decision on basis of diverse stress and metabolic inputs, regarding whether to self-renew, differentiate or die. In addition G1 and G2 GAPS allow the repair of DNA damage and replication errors. Only specialized cells, such as those that populate the hematopoietic system or line the gut epithelium maintain active proliferation. All cells have the capacity to enter quiescence and all quiescent cells, with the exception of those that have reached a state of terminal differentiation, have the capacity to re-enter the cycle (Malumbres and Barbacid 2001; Massague 2004).

Normal cells only proliferate when they receive the correct instructions from other neighboring cells, this collaboration results in a tissue size adequate to the body needs. But DNA contained in every mammalian cell is under constant attack by damaging agents and as a consequence cells can acquire mutations. Among the mechanisms that cells have developed to cope with this attack on their genetic material are DNA repair processes. In addition, cells respond to DNA damage by halting cell cycle progression, cell cycle checkpoints are mechanisms by which the cell actively halts progression through the cell cycle until it can ensure that an earlier process, such as DNA replication is completed. If the DNA damage cannot be repair, cells can respond with other mechanisms: either undergo programmed cell death (Kastan and Bartek 2004) or enter into a permanent withdrawal form the cell cycle, cellular senescence (Campisi 2005).

Cancer results when mutated cells survive and proliferate inappropriately disrupting the existing proliferation-apoptosis balance. The uncontrolled cell growth is often accompanied with the invasion of surrounding tissue and the spread of malignant cells. Several stimuli like U.V.-radiation, pro-carcinogenic compounds or infection by viruses are known to increase the risk for developing aberrant cell growth.

In the last century our knowledge of the molecular mechanisms leading to cancer has greatly improved. The discovery that the evolution of human cells into malignant derivates is driven by the aberrant function of genes was one of the most important milestones in cancer research. This finally led to the concept that some genes have a fundamental role as positive (oncogenes) or negative (tumor suppressor genes) regulators of normal cell proliferation. Activating mutations in oncogenes and loss of function mutations in tumor suppressor genes are leading to cancer aberrant proliferation. Not only changes in protein structure but also changes in the transcription are responsible for these activation or loss of function mutations. Increase of transcription can activate oncogenes or tumor suppressor genes, in contrast loss of transcription by methylation associated silencing can inactivate certain tumor suppressors.

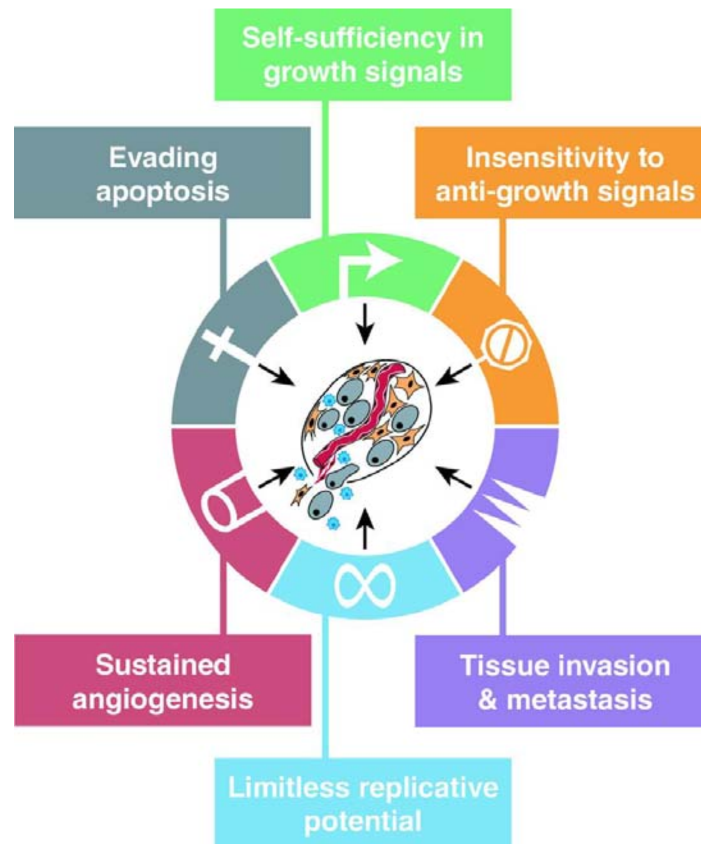


Figure 1. The hallmarks of cancer. Tumor cells have acquired through different mechanisms six essential alterations, which are depicted in the figure. These alterations result in uncontrolled cell proliferation, invasion of surrounding tissue and spread of malignant cells. Figure from Hanahan and Weinberg 2000.

For a tumor to develop there have to be mutations in a dozen or more genes that control proliferation (Weinberg 1996). Features acquired by the malignant cell have been recently

defined by Hanahan and Weinberg as the manifestation of six essential alterations in normal cell physiology: self sufficiency in growth signals, insensitivity to growth inhibitory signals, evasion of programmed cell death (apoptosis), limitless replicative potential, sustained angiogenesis and tissue invasion and metastasis (figure 1). Each of these physiological changes represents the successful breaching of an anticancer defense mechanism (Hanahan and Weinberg 2000).

Despite all the advances in oncology there is not an efficient therapy that highly specific destroy all kind of cancer cells. New knowledge in the molecular mechanisms leading to cancer should provide insights how new anticancer strategies can be devised.

1.2. A link between inflammation and cancer

Up to now most of the attention in the treatment of cancer has focus on destroying the malignant cell paying little attention to the cancer microenvironment (Karin and Greten 2005); but the pre-malignant cells are not an isolated island rather a focus of intense tissue interactions (Weiss 1971). Among these interactions the response of the body to cancer which has many parallels with inflammation and wound healing has received special attention in the last years. The importance of this inflammatory response has been highlighted by Balkwill and Mantovani who made a remarkable comparison: if genetic damage is “the match that lights the fire” of cancer, some types of inflammation may provide the “fuel that feds the flame” (Balkwill and Mantovani 2001).

Inflammation is the body's reaction to invasion by infectious agent, antigen challenge or even just physical damage. This biological response involves complex interactions between different cell types, the biological mediators include cytokines, which promote cell activation and proliferation, and chemokines which induce chemotaxis and cell migration.

Inflammation and cancer association has been already suggested in the 19th century by Virchow who noted leukocytes in the neoplastic tissues (Balkwill and Mantovani 2001), later Harold Dvorak referred to this association by comparing tumors to wounds that never heal (Dvorak 1986). Epidemiological studies supported these observations and showed that susceptibility to cancer increases when tissues are chronically inflamed. Indeed, long term use of non-steroidal drugs like aspirin, as well as natural compounds like cucurmin, green tee

extract, ginseng extract or resverastol (polyphenol from red wine) reduces the risk of several cancers in organs susceptible to chronic inflammation (colon, lung, esophagus and ovaries, as well as Hodgkin's lymphoma; Gupta and Dubois 2001; Aggarwal, Kumar et al. 2003; Garber 2004). However, this association is more complicated than portrayed, as illustrated by the fact that, although non-steroidal drugs decrease the risk for certain cancers, they might increase the risk for others (Cerhan, Anderson et al. 2003; Chan, Giovannucci et al. 2004; Schernhammer, Kang et al. 2004).

Tumorigenesis is a multistep process and inflammation seems to be associated with each step. During tumor initiation the DNA of the cell is mutated spontaneously or by chemical or physical carcinogens, leading to the activation of oncogenes or the suppression/inactivation of tumor suppressor genes. The role of inflammation in this early step of tumor formation seems to be not a general one and might be restricted to cancer that arises after massive cell destruction.

Tumor promotion is characterized by the clonal expansion of initiated cells, owing to increased cell proliferation and/or reduced cell death. Inflammation seems to be an important factor during this phase of tumor development; in fact several pro-inflammatory cytokines are known to promote tumor growth such as tumor necrosis factor (TNFα), interleukin 1 (IL-1), interleukin 6 (IL-6) or interleukin 8 (IL-8) (Pikarsky, Porat et al. 2004; Karin and Greten 2005). In addition, the link between activating mutations in oncogenes and inflammation has been indicated in a recent study which showed that activation of *Ras* proto-oncogenes in cancer results in up-regulation of the inflammatory cytokine IL-8 which acts as a chemokine and in turn promotes tumor associated inflammation, angiogenesis and eventually tumor growth (Sparmann and Bar-Sagi 2004; Karin 2005).

Increased tumor size as well as invasion and metastasis, are the ultimate characteristics of malignant tumor progression. In addition of being immunosuppressive, infiltrating leukocytes might contribute to these processes by producing metalloproteinases (MMPs) as well as growth and angiogenic factors.

Two cytokines are of outstanding importance in the modulation of the immune response. Tumor necrosis factor alpha (TNFα) plays a key role in supporting the innate immune response by promoting innate cell stimulation (macrophages, dentritic cells, natural killer

cells, neutrophils) and pro-inflammatory cytokine secretion. In contrast, TNF-related weak inducer of apoptosis (TWEAK, also known as Apo3L or TNF12) has recently been identified as crucial for curtailing the innate response. Whereas TNFα activates transcription of immunostimulatory genes by promoting STAT-1 activation and the association of the NF-κB subunit p65 with the transcriptional co-activator p300, TWEAK represses STAT-1 activity and induces binding of p65 to histone deacetylase 1 (HDAC-1), inhibiting pro-inflammatory gene transcription. TWEAK's function may have evolved to guard against the development of potentially harmful excessive inflammatory and autoimmune responses (Maecker, Varfolomeev et al. 2005).

The molecular mechanisms connecting tumorigenesis to chronic inflammation remained largely unresolved until two studies showed that the molecule linking cancer and inflammation is the nuclear factor kappa B (NF-κB) (Greten, Eckmann et al. 2004; Pikarsky, Porat et al. 2004).

1.3. NF-κB: a major player in cancer progression

1.3.1. The NF-κB transcription family

NF-κB is not a single protein but a family of closely related factors that associate in homo- or heterodimers, which bind a common sequence motif known as the κB site. The NF-κB family is best known for its functions in immune, inflammatory and acute phase responses, but also plays important roles in cell growth, apoptosis and oncogenesis (Kucharczak, Simmons et al. 2003).

NF-κB proteins can be classified in two subgroups: one consists of c-Rel, RelA (p65) and RelB. These proteins share a N-terminal Rel-homology domain (RHD) which includes a leucine zipper dimerization motif that enables them to translocate to the nucleus, form dimers and bind to the κB DNA sites (5'-GGGRNNYYCC-3'). The second class consists of NF-κB1 (p105/p50) and NF-κB2 (p100/p52) which are synthesized as precursors (p105 and p100) with an N-terminal RHD and a C-terminal series of ankyrin repeats, the C-terminal portions of p150 and p100 prevent nuclear entry and are removed by ubiquitin dependent degradation giving the mature forms. The most abundant form of NF-κB in cells is a heterodimer of p65

and p50 that is sequestered in the cytoplasm through association with several IκB factors and as a consequence is inactive. There are two major activation pathways for NF-κB complexes: the canonical or classical pathway is triggered by viral and microbial infections or exposure to pro-inflammatory cytokines; the non-canonical or alternative pathway is activated in response to other stimuli such as lymphotoxin B (LTβ) and B cell activating factor belonging to the TNF family (BAFF) (figure 2).

The classical activation pathway applies to dimers that are composed of c-Rel, Rel A (p65) and p50 which are held inactive in the cytoplasm by IκB. A more detailed view of the IκB-NF-κB complex reveals that the IκB proteins mask only the nuclear localization sequence of p65 whereas the nuclear localization signal of p50 is accessible. This sequence in combination with the nuclear export sequences present in IκB and p65 results in constant shuttling of IκBα-NF-κB between the nucleus and the cytoplasm with a net accumulation in the cytoplasm (Huang, Kudo et al. 2000). Upon activation IκBα is degraded, altering this balance, which results in predominant nuclear localization. IκBα degradation is a tightly regulated process that is initiated upon specific phosphorylation by activated inhibitors of κB kinases (IKKs) (Karin and Ben-Neriah 2000).

The non-canonical or alternative pathway affects NF-κB2 which preferently dimerizes with RelB. This pathway selectively activates NF-κB inducing kinase (NIK) and IKKα triggering phosphorylation and ubiquitination of p100. This ubiquitination induces proteolytic removal of the C-terminal domain of p100, allowing its translocation to the nucleus. p100 processing is tightly regulated with only minimal processing in unstimulated cells. Unlike p100, p105 (p50) processing is not as strictly regulated and as a consequence p50 homodimers are present in the nuclei of unstimulated cells. These homodimers can associate with HDAC-1 and repress the transcription from κB dependent promoters. This p50 dependent HDAC-1 recruitment to NF-κB responsive gene promoters is not a general phenomenon as shown by constitutively or rapidly NF-κB activated genes which have high levels of associated acetylated histone H4 in resting cells. Once in the nucleus, heterodimeric NF-κB complexes bind to most κB sites with greater affinity than the p50 homodimers and can displace them.

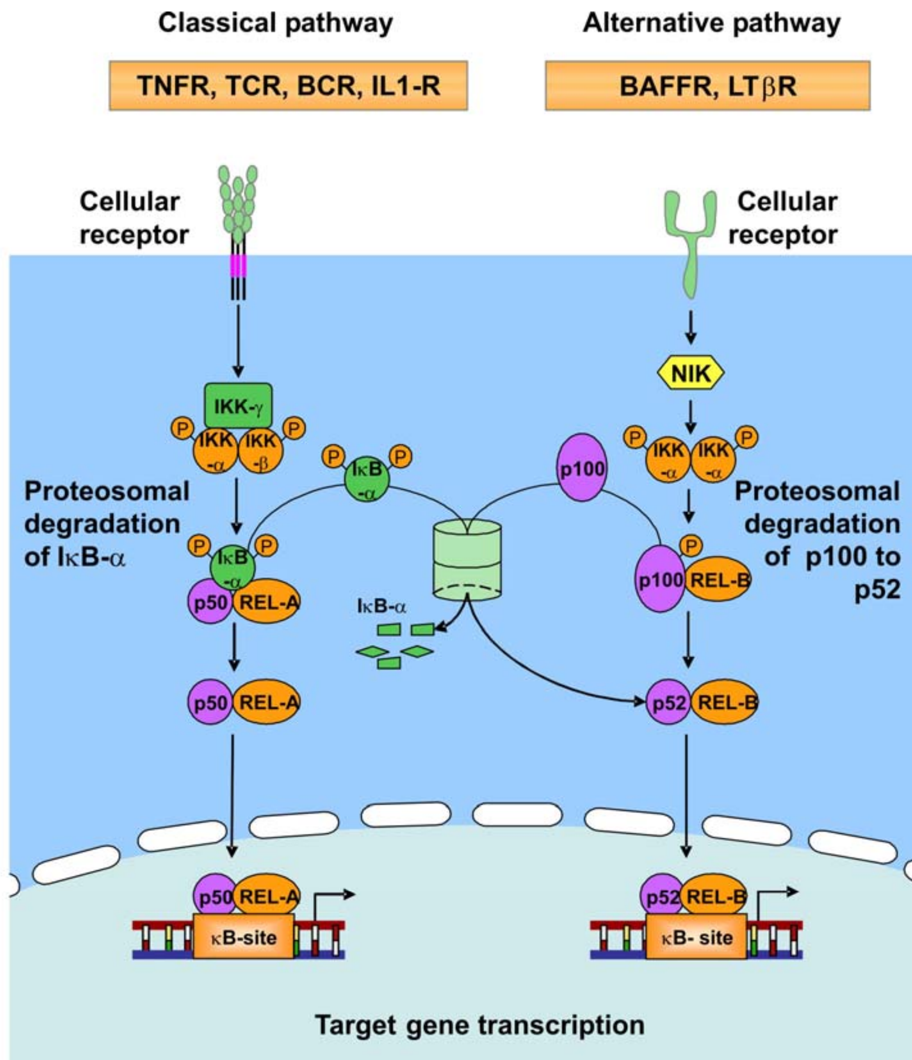


Figure 2. NF-κB activation pathways. The members of the NF-κB transcription factor family can be activated through two pathways triggered by different stimuli (see text for explanations). Adapted from Karin and Greten 2005. BAFFR, BAFF (B cell activating factor belonging to the TNF family receptor) receptor; BCR, B-cell receptor; IL-1R, interleukin 1 receptor; IκBα, inhibitor of κBα NF-κB, nuclear factor kappa B; IKK α/β/γ, inhibitor of NF-κB (IκB) kinase α/β/γ; LTβR, lymphotoxin B receptor; NIK, NF-κB inducing kinase; p100, p52, p50, REL-A, REL-B, NF-κB subunits; TCR, T-cell receptor; TNFR, TNFα receptor.

Each NF-κB dimer is likely to have distinct regulatory functions, however many of the target genes are common to several NF-κB factors. Genes regulated by NF-κB belong to different functional classes: negative feedback (*inhibitor of κBα* [*IκBα*], *inhibitor of κBβ* [*IκBβ*]), immunity (chemokines, cytokines), anti-apoptosis (*B-cell lymphoma* X_L [*BCL-*X_L], cellular *inhibitors of apoptosis* [*c-IAPs*]) and proliferation (*cyclin D1*, *c-myc*); most of these genes can contribute to the alterations observed in the malignant cell (Mayo and Baldwin 2000).

1.3.2. NF-κB as the link between cancer and inflammation

Recent experimental evidence from two mouse models has led to the conclusion that NF-κB is a central molecule linking inflammation and tumorigenesis (reviewed by Balkwill and Coussens 2004; Karin and Greten 2005).

Multidrug resistance 2 gene (*MDR2*) knockout mice provide a model for cholangitis (bile duct inflammation) caused by bile acid and phospholipid accumulation which leads to the appearance of hepato-carcinogenesis between 7 and 14 months after birth. Inactivation of NF-κB, through the expression of a non-degradable IκBα variant (IκB super-repressor) under the control of a promoter that is highly active in hepatocytes, blocked tumor development. These tumor suppressive effects of NF-κB inhibition were associated to apoptosis of the hepatocytes. However, if the inhibition was only during the first 7 months, that is during the initiation and early promotion phases of tumorigenesis, non-blocking effect was detected. Hence, it seems that whereas NF-κB is dispensable for the early pre-malignant phase (tumor initiation) is essential for subsequent tumor promotion. A more detailed analysis of this knockout revealed an up-regulation in TNFα, a cytokine which is known to activate NF-κB, in the non-hepatocyte fraction of the liver. Indeed when mice were treated with a neutralizing antibody against TNFα, apoptosis of the hepatocytes was induced as effective as inhibiting NF-κB between 7-14 months which indicates that the activation of NF-κB is through the pro-inflammatory cytokine TNFα (Pikarsky, Porat et al. 2004).

The tumor promoting activity of NF-κB has been also shown in a mouse model of colitis-associated cancer where NF-κB activation suppressed the apoptosis of chemically transformed pre-malignant cells. In this model, mice were injected first with a pro-carcinogen, followed by oral administration of dextran-sulphate sodium salt (DSS) which induced chronic colitis; through disruption of the intestinal barrier and exposure of macrophages in the lamina propia to enteric bacteria. This exposure resulted in the activation of NF-κB in macrophages leading to the production and secretion of pro-inflammatory cytokines which activated NF-κB in the epithelial cells. Enterocyte ablation of IKK-β blocked NF-κB activation and decreased tumor incidence indicating that the NF-κB pathway was necessary in this case for early tumor promotion. In contrast to the *MDR2* knockout, most of the apoptotic cell death of IKK-β deficient enterocytes occurred within a few days of exposure to carcinogen and DSS, most

probably due to a failure in the induction of anti-apoptotic genes in enterocytes. A second mechanism through which NF-κB affected tumor promotion resulted from the transcriptional induction of pro-inflammatory factors by myeloid cells. In agreement with this observation, ablation of IKK-β in myeloid cells decreased not only tumor number but also tumor size. The decrease in tumor size was a consequence of the diminished proliferation of transformed epithelial cells which required growth factors (like IL-6) produced by the myeloid cells (Greten, Eckmann et al. 2004).

All these evidences point to a role for NF-κB in inflammation associated tumorigenesis (figure 3). In the malignant cell NF-κB activated by oncogenes or inflammatory cytokines functions as a transcription factor for anti-apoptotic and proliferative genes. This proliferation effect is promoted by the leukocytes present in the tumor where NF-κB induces the production of cytokines, either inflammatory or proliferative, which then act on the tumor cells. The conclusion of these models for cancer therapy is innovative, whereas eradicating the primary cause of tumorigenesis is currently unattainable, disrupting the signaling context of the evolving tumor could be a more realistic objective. Intermittent suppression of a major signaling factor could be a tool to inhibit tumor progression in chronic inflammatory diseases associated with high cancer risk.

Intriguingly, chronic inflammations do not always predispose to cancer. One such prominent exception is psoriasis, a chronic cutaneous inflammatory disease which is seldom accompanied by cancer. This particular inflammatory condition has been recently reviewed by Nickoloff, keratinocytes exposed to the cytokines present in the psoriatic plaque become growth arrested, resistant to apoptosis and present several senescence markers (Nickoloff, Ben-Neriah et al. 2005). There are other features in addition to the senescence state that explain the absence of cancer in the psoriatic plaques; secreted proteins like maspin and transforming growth factor β (TGF-β), overproduced by the senescent keratinocytes potentially mediate an anti-tumorigenic response (Tremain, Marko et al. 2000; Nickoloff, Lingen et al. 2004).

These evidences indicate that inflammatory reactions have dualistic and perhaps even opposing influences depending on the cellular context, strength-persistence of the signals and other microenvironmental factors (Nickoloff, Ben-Neriah et al. 2005 and references herein).

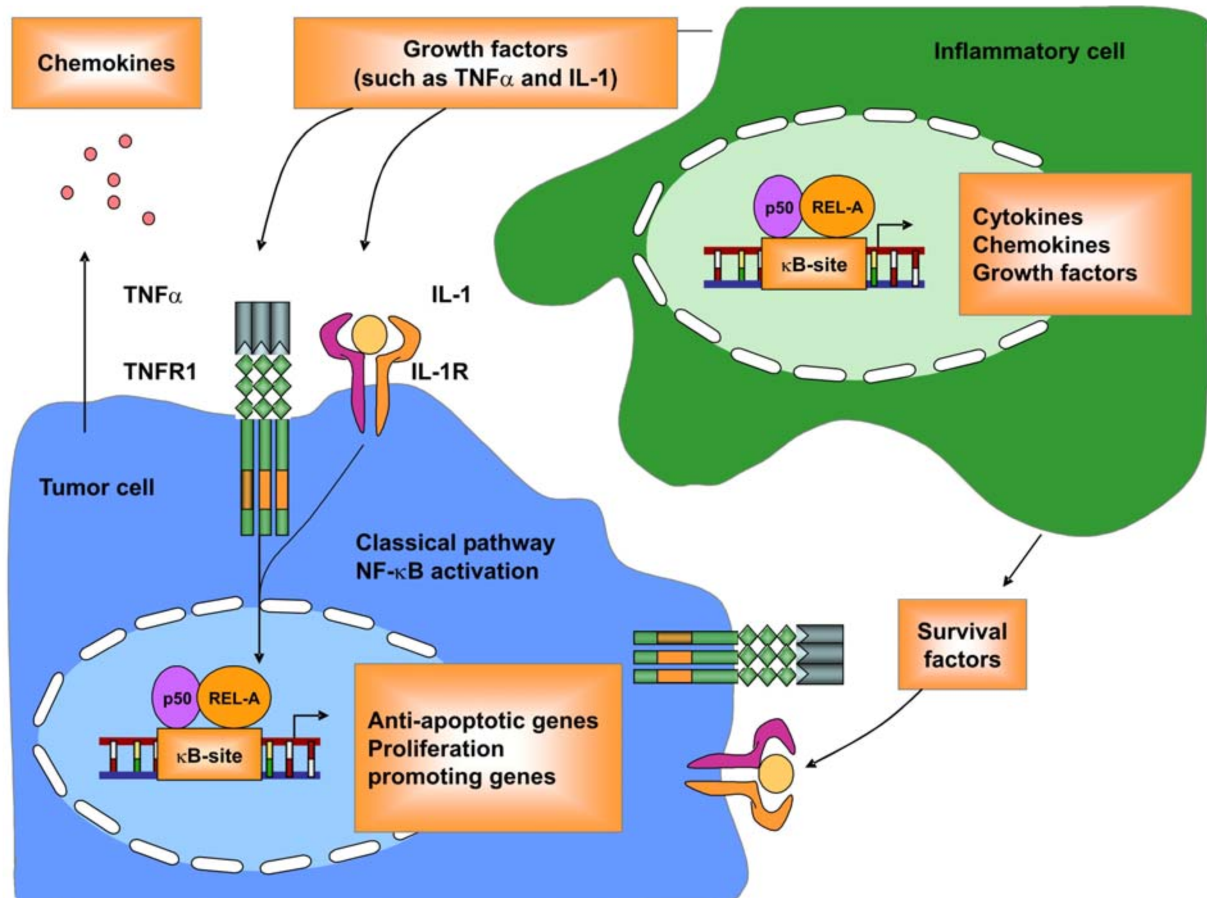


Figure 3. Functions of NF-κB in inflammation associated tumorigenesis. In inflammatory cells NF-κB induces the production of cytokines, growth and survival factors. In the tumor cell, the cytokines produced by the inflammatory cell activate NF-κB leading to the up-regulation of anti-apoptotic genes and secretion of chemokines which attract more inflammatory cells. Adapted from Karin and Greten 2005. IL-1, interkeukin 1; IL1-R, interleukin 1 receptor; NF-κB, nuclear factor kappa B; p50, RELA, NF-κB subunits; TNFα, tumor necrosis factor α; TNFR1, TNFα receptor.

1.3.3. NF-κB activation in tumors of different origins

The first link between carcinogenesis and NF-κB was afforded by the identification of the p50 NF-κB subunit as a Rel protein, since a viral member of this family (v-Rel) is the oncoprotein of the reticuloentotheliosis (REV-T) retrovirus which causes aggressive lymphomas in animal models. Additional evidence was provided by the interaction of other, unrelated viral oncoproteins, with the IKK complex which caused constitutive IKK and NF-κB activation. One example is TAX of the human *T-cell leukemia virus* (*HTLV-1*). Activation of NF-κB is also observed in the infection caused by herpes virus *Epstein-Barr* virus (*EBV*) which is

implicated in Burkitt's lymphoma or the herpes virus 8 linked to Kaposi's sarcoma (Kucharczak, Simmons et al. 2003). NF-κB can also be induced by cellular oncogenes, such as *Ras* (Finco, Westwick et al. 1997) or *BCR-ABL* (Hamdane, David-Cordonnier et al. 1997). Chromosomal translocations, amplifications, deletions and mutations that affect members of the NF-κB or IκB protein families can be found in diverse hematological diseases. Amplification of c-Rel is observed in non-Hodgkin's B-cell lymphomas, chromosomal rearrangements in the NF-κB2 locus can be detected in B- or T-cell lymphomas and the p-100 NF-κB2 precursor (p100) is over-expressed in some breast cancer cell lines and tumors (Sovak, Bellas et al. 1997; Cogswell, Guttridge et al. 2000).

Moreover, a mouse model of liver cancer has been used to demonstrate the direct association between NF-κB and chemically induced tumorigenesis in non-inflammatory conditions. NF-κB activation was prevented by ablation of the gene encoding IKKβ in hepatocytes. Cancer was induced by the administration of pro-carcinogen, dietylnitrosamine (DEN) which induced hepatocyte DNA damage and mutations as well as hepatocyte death, to 2 week old mice. Pro-carcinogen treatment of these mice resulted in increased cell death by either apoptosis or necrosis of hepatocytes. Thus, deletion of IKKβ should greatly increase the sensitivity of these cells to death inducing challenges, including the pro-carcinogen. However, upon DEN administration the animals showed increases in tumor size, growth rate and aggressiveness compared to control mice. This was attributed to the release of cytokines from Kupffer cells (macrophages in the liver) stimulated by the increased necrotic liver injury. These cytokines in turn stimulated hepatocyte growth factor (HGF)-production in stellate cells, which induced cell growth and hence resulted in an increased tumor size. In accordance with this observation, simultaneous inhibition of the IKKβ in both Kupffer cells and hepatocytes resulted in fewer and smaller hepatocarcinomas. These findings also raise the possibility that tumor necrosis might be responsible for the recruitment and activation of myeloid cells in other solid tumors (Maeda, Kamata et al. 2005).

NF-κB not always performed a tumor promotion function in all the cellular contexts and with all the variety of stimuli. When is activated by certain inducers or when particular tumor suppressor proteins are active, RelA represses rather than activates anti-apoptotic gene expression (Perkins 2004). An example of such activation is provided by the tumor suppressor

gene *ADP-ribosylation factor* (*ARF*). ARF has been reported to induce the association of RelA with HDAC-1, thereby turning it into a repressor of gene expression without affecting NF-κB binding activity. In the absence of ARF, the BCR-ABL oncogene stimulates NF-κB activity whereas in its presence a repression of transcription was seen. ARF-modulated NF-κB repressed also the transcription of BCL-X_L which resulted in increased sensitivity to other apoptotic stimuli such as TNFα or ectoposide.

Recently Campbell and co-workers have reported that certain cancer drugs (daunorobicin) hijack the RelA subunit of NF-κB and turn it into dominant negative repressor of survival gene expression. The authors indicated that a change in the phosphorylation status of NF-κB could be responsible for this switch; indeed U.V.-C and daunorobicin induced the association of RelA with HDAC and resulted in deacetylation of histones at RelA target promoters (Campbell, Rocha et al. 2004; Miyamoto 2004).

1.4. A genomewide screen for recovery of survival genes

1.4.1 TNFα as an inducer of NF-κB signaling

TNFα is a pro-inflammatory cytokine produced in many cancers and its presence is generally associated with poor prognosis. Indeed, this cytokine is one of the strongest inducers of NF-κB gene transcription in the malignant and inflammatory cells; TNFα chronically produced in the tumor microenvironment enhances development and spread of malignant cells as has been highlighted in the other sections. However TNFα is also a crucial effectors molecule in CD8 and NK-cell mediated killing of some tumor cells. In addition, at supra-physiologic levels, this cytokine has impressive tumor destructive properties (Balkwill 2002). The causes of these paradoxical actions have the origin in the TNFα signaling pathway.

TNFα is a 17 kDa polypeptide which binds to its receptor in a trimeric form; there are two receptors for this cytokine: tumor necrosis factor receptor one (TNFR1) which is expressed ubiquitously and tumor necrosis factor receptor two (TNFR2) which is expressed in endothelial cells and cells of the immune system. The latter receptor can only be fully activated by the membrane bound form but not by soluble TNFα (Grell, Douni et al. 1995).

TNFα is mainly produced by macrophages but is also synthesized by other tissues including lymphoid cells, mast cells, endothelial cells, fibroblast and neuronal tissue.

TNFR1 signaling proceeds via the sequential formation of two distinct complexes (figure 4). After binding of TNFα to its receptor TNRF1, the silencer of death domain protein (SODD) dissociates from TNFR1 and the later recruits the adapter protein TNFR-associated death domain (TRADD). This activated receptor then serves as an assembly platform for binding of several molecules: TNFR-associated factor-2 (TRAF2) and receptor-interacting protein (RIP) (Wajant, Pfizenmaier et al. 2003). This first complex stimulates pathways leading to the activation of NF-κB (canonical pathway) and JNK. JNK activation has pro- or anti- apoptotic effects depending on the duration of the stimulus; an acute activation is anti-apoptotic while a chronic activation has the opposite effect. A second complex, death-inducing signaling complex (DISC) which lacks TNFR1 but includes Fas-associated death domain (FADD) and pro-caspases-8 and -10 is subsequently formed in the cytoplasm by modification of complex I. This secondary complex (complex II) initiates the apoptotic response by activation of the pro-caspases (Barnhart and Peter 2003; Micheau and Tschopp 2003).

In contrast to complex I, the composition of complex II in apoptosis resistant and sensitive cells differs. In resistant cells, complex II comprises increased amounts of the anti-apoptotic proteins c-IAP1 and FLIP, the expression of which is regulated by the transcriptional activity of NF-κB (Wang, Mayo et al. 1998; Micheau, Lens et al. 2001). FLIP (Flice-inhibitory protein) availability at the moment complex II is formed is dependent on a signal triggered by complex I. If NF-κB activation promotes the expression of FLIP the pro-apoptotic activity of caspase-8 is inhibited.

The interesting implication of this model is the prediction that the decision between life and death in the TNFα signaling pathway is controlled by at least one checkpoint. This checkpoint is triggered immediately after TNFR1 engagement but is operational only a few hours later, at a point when the success of the transcriptional activity of NF-κB can be assessed. Cells defective in NF-κB signaling, will have low quantities of anti-apoptotic proteins and will be eliminated through apoptosis (Micheau and Tschopp 2003).

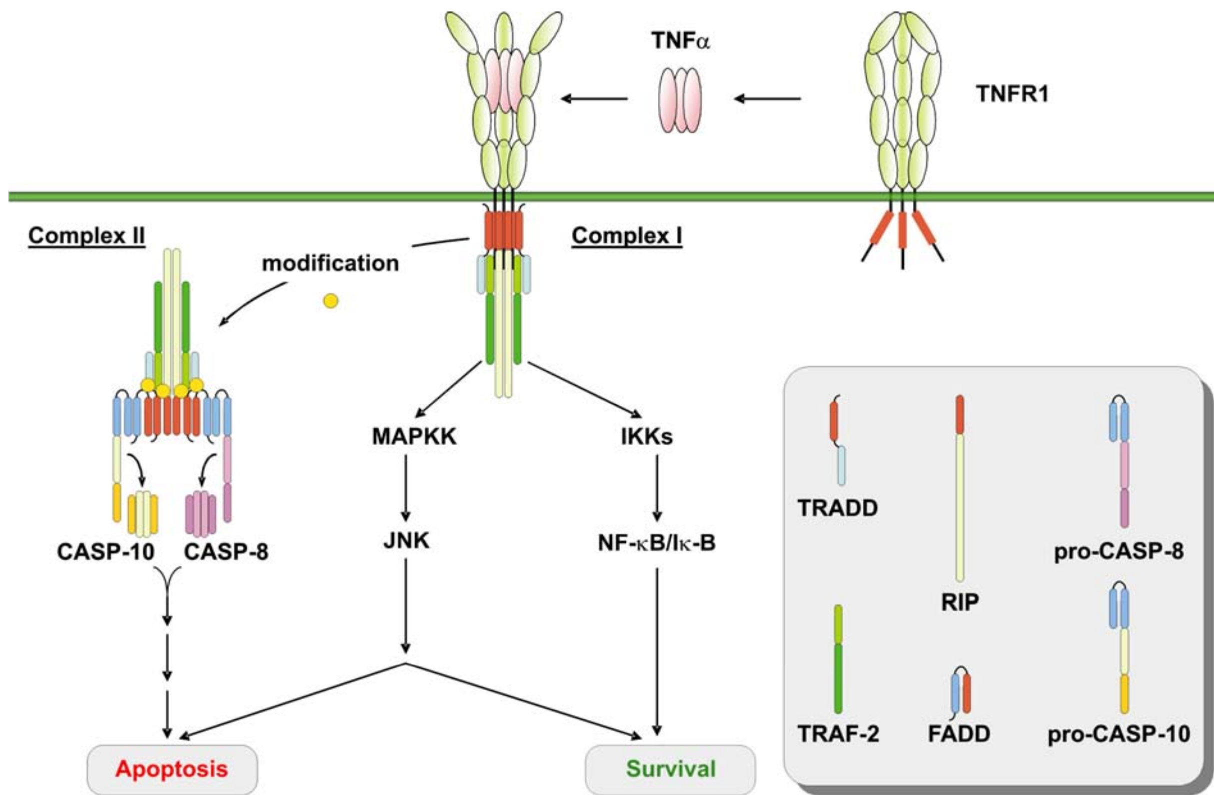


Figure 4. TNFα signaling pathways. TNFα can activate a cell survival pathway and an apoptotic response through two different signaling complexes. Adapted from Barnhart and Peter 2003, see text for explanations. CASP-8, caspase 8; CASP-10, caspase 10; IKK, IκB kinase complex; FADD, Fas-associated death domain; JNK, c-Jun N-terminal kinase; RIP, receptor interactive protein; MAPKK, mitogen-activated protein kinase kinase, NF-κB, nuclear factor kappa B; TNFα, tumor necrosis factor α; TNFR1, TNFα receptor 1; TRADD, TNF receptor associated death domain; TRAF2, TNF receptor-associated factor-2; pro-CASP denotes the inactive pro-caspases.

The response to TNFα is variable and depends on the physiological context, thereby is important to understand the molecular basis on which the cell chooses between life and death (Kyriakis 2001). In type I cells, active caspase 8 recruited by complex II is sufficient to induce caspase 3 activity and apoptosis. However in type II cells, caspase 8 is not able to induce caspase 3 and therefore the apoptotic process depends on a mitochondrial amplification loop. Small amounts of caspase 8 are able to activate the protein BID which in turns translocate to the mitochondria and promotes the release of cytochrome c and Smac/Diablo. The release of cytochrome from the mitochondria allows the formation of the apoptosome consisting of cytochrome c, APAF-1 and caspase 9 (figure 5), which then

triggers apoptosis. In addition, Smac/Diablo binds and antagonizes the caspase inhibitors X-chromosome-linked IAP (XIAP), cIAP1 and cIAP2 (Wajant, Pfizenmaier et al. 2003).

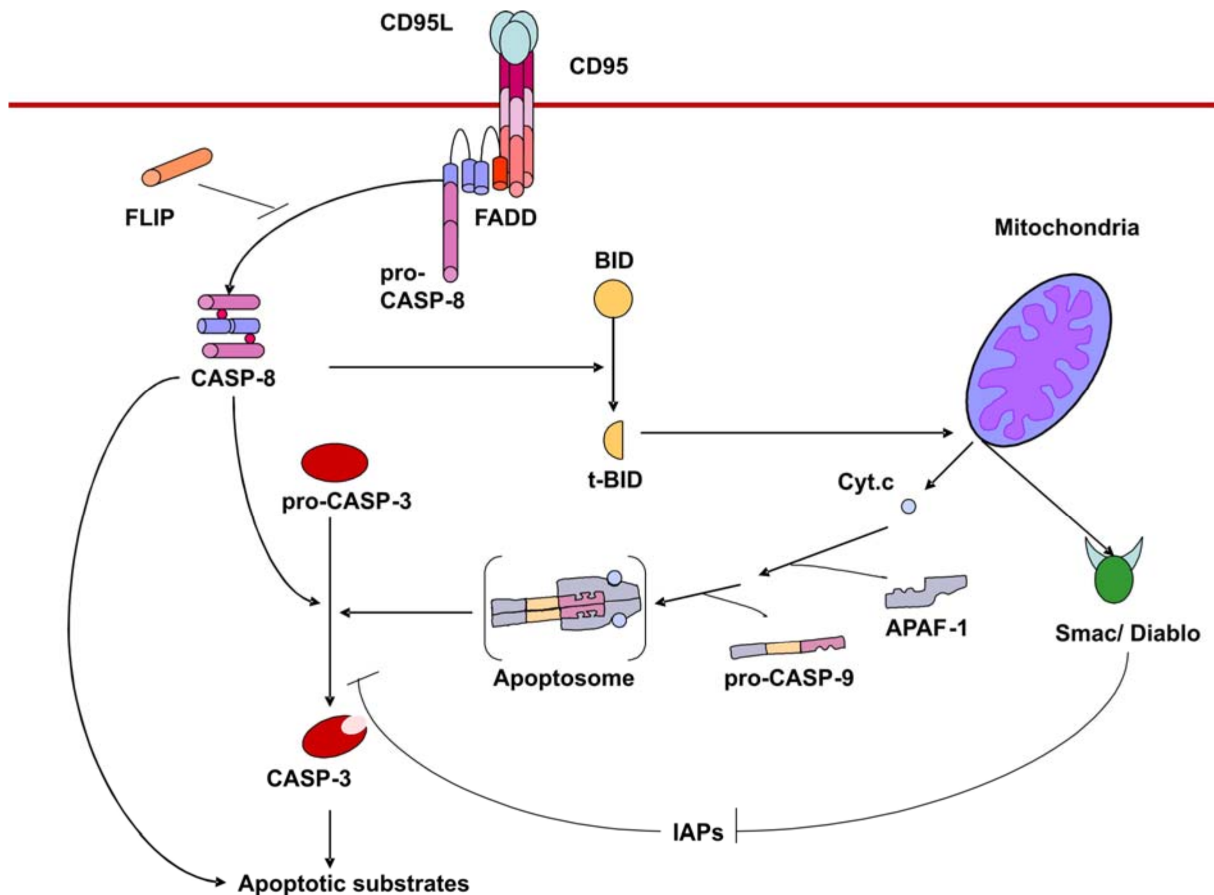


Figure 5. Crosstalk between the mitochondrial and receptor apoptotic pathways in type II cells. Adapted from Hengartner 2000; see text for explanation. APAF-1, apoptotic protease activating factor; BID, BH3 interacting domain death agonist;CASP-8, caspase 8; CASP-3, caspase 3; Cyt.c, cytochrome c; FLIP, Flice-inhibitory protein; IAPs, inhibitor of apoptosis proteins; pro-CASP denotes the inactive pro-caspases; Smac/ Diablo, second mitochondria-derived activator of caspases /direct IAP-binding protein with low pI.

There are several mechanisms in addition to the activation of survival genes through which NF-κB protects cells against apoptosis. Examples include the transcriptional activation of Gadd45β (De Smaele, Zazzeroni et al. 2001), XIAP (Tang, Minemoto et al. 2001) and proteins which inhibit reactive oxygen species (ROS) accumulation (Sakon, Xue et al. 2003; Kamata, Honda et al. 2005) and thus protect cells from a permanent activation of JNK which would lead them to TNFα induced apoptosis.

1.4.2. Combination of gene trap with Cre/*loxP* induced recombination allows detection of transiently TNFα–induced genes

TNFR1 induces apoptosis in a variety of cell lines; therefore activation of survival genes in these cells is a transient event. Identification of these genes is from special interest for cancer research, as permanent expression of survival genes is responsible for the resistance to apoptosis which contribute to the cancer phenotype. These genes would provide new targets for diagnostic and/or therapeutic.

The identification of transiently induced genes can be achieved by a gene trap, gene traps are plasmid or retrovirus-based vectors containing a reporter gene which are introduced into a random collection of chromosomal sites. Gene traps have evolved from enhancer-trap vectors, a molecular tool used to identify and characterize mammalian enhancer sequences from cell lines (Weber, de Villiers et al. 1984; Stanford, Cohn et al. 2001), to promoter-, poly A- and gene trap vectors. Specifically, the one used in this work, the promoter-trap vector consists of a promoter-less reporter gene and a selectable marker. Reporter expression occurs when the vector inserts into an exon to generate a fusion transcript that comprises the upstream endogenous exonic sequence and the reporter gene (von Melchner and Ruley 1989).

However, the activation of the gene trap has to be coupled to the TNFα treatment; this can be achieved by conditional recombinase systems that allow gene expression to be abrogated in a temporally and spatially controlled manner. The promoter gene trap containing a recombinase has to be combined with a switch reporter system.

The promoter gene trap might contain *Cre* as selectable marker, Cre recombinase of the bacteriophage P1 belongs to the integrase family of site-specific recombinases (Nagy 2000) and only few molecules are needed to induce recombination. Cre is a 38 kDa protein that catalyzes the recombination between 2 of its recognition sites, called *loxP* (Hamilton and Abremski 1984). The *loxP* consensus sequence consists of a core spacer sequence of 8 bp and two 13 bp palindromic flanking sequences which define the orientation of the *loxP* sites. A single recombinase molecule binds to each palindromic half of a *loxP* site, and then the recombinase molecules form a tetramer, bringing the two *loxP* sites together (Voziyanov, Pathania et al. 1999). The recombination is permanent and occurs between the spacer area of

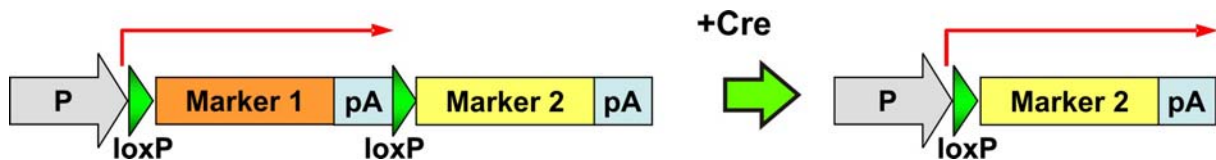
the *loxP* sites, the post-recombination *loxP* site is formed from the two complementary halves of the pre-recombination sites and the spacer area between the *loxP* sites is discarded.

The switch reporter system contains a constitutively active promoter upstream of two transcription units, each of which consists of a coding region with a polyadenylation signal. The 5'-cassette, encoding a protein conferring resistance against an antibiotic is flanked by tandemly arranged *loxP* sites. In the starting configuration only this cassette is transcribed, making cells with a stable integration of the switch resistant against an appropriate antibiotic. When Cre is present, the floxed DNA region is excised leading to transcriptional activation of the downstream cassette.

The promoter trap U3Cre combined with Cre/*loxP* site specific recombination in a reporter construct has permitted the identification of genes specifically induced after different biological stimuli: growth factor deprivation (Russ, Friedel et al. 1996; Wempe, Yang et al. 2001) or glucocorticoids (Wan and Nordeen 2002). This strategy has been used in the present thesis for the recovery of genes specifically induced by TNFα in the MCF-7 breast cancer cell line. TNFα induced apoptosis in MCF-7 cells was recovered upon Cre induced recombination, which transformed a transient activation of the gene trap insertion into a permanent switch in the reporter construct. This switch allowed the expression of the second marker gene (*dnFADD*) which blocks the apoptotic branch of the TNFα signaling (figure 6).

Several features of this strategy made it suited for the isolation of genes transcriptionally regulated by TNFα in a process leading to apoptosis. First, permanent selectable switching between two selectable marker genes protects cells against apoptosis. Second, by uncoupling the expression of the trapped cellular gene from the expression of the selectable marker gene, the strategy allows the identification of genes that are even transiently expressed. Third, this technique is very sensitive, only few molecules of Cre are needed to permit recombination of the reporter vector allowing the identification of genes which are even at extremely low levels. Fourth, this technique is a genome wide scanning which might allow the identification of new genes in the human genome.

A.-



B.-

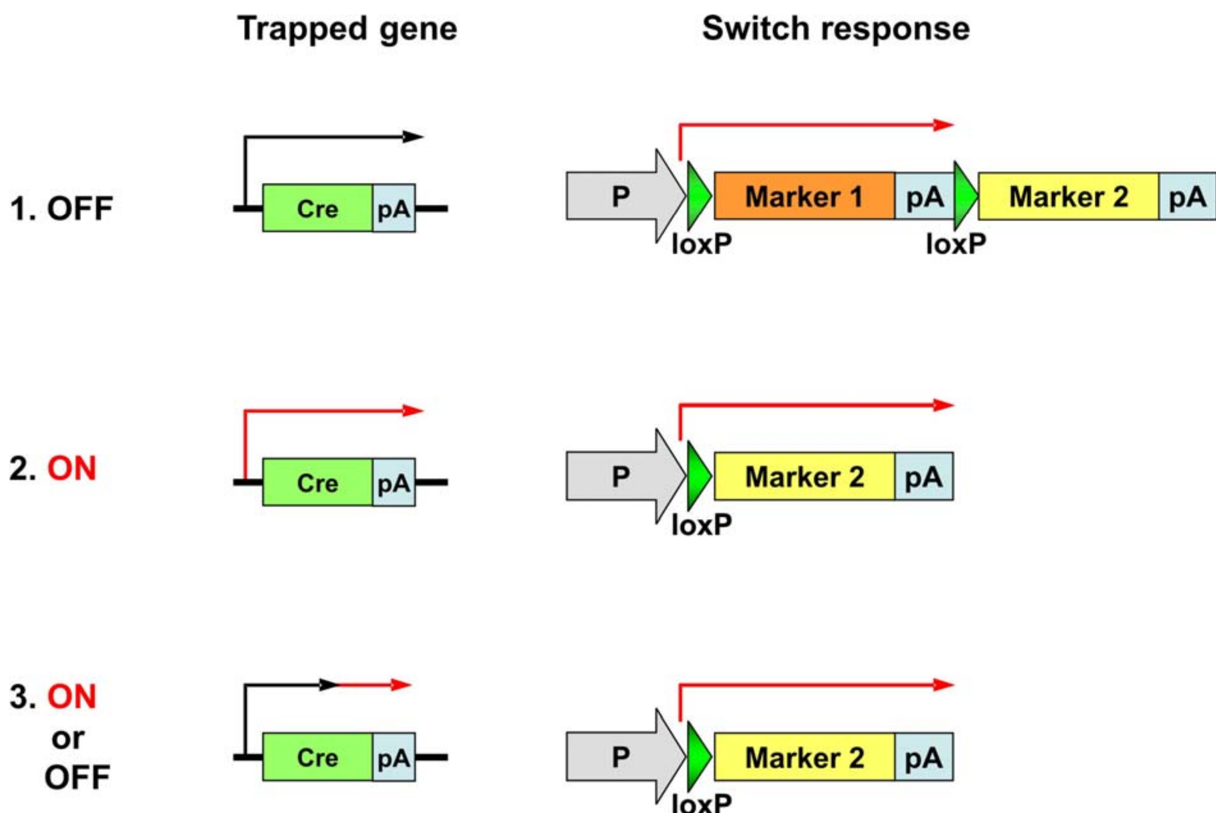


Figure 6. Strategy for the recovery of inducible genes. A.- Anatomy of the switch cassette. A selectable marker gene (marker 1), flanked by *loxP* sites (green triangles) in direct orientation relative to each other, is expressed from a constitutively active promoter (P) and simultaneously blocks the expression of a second, downstream marker gene (marker 2) by premature polyadenylation (pA). Cre deletes the upstream marker cassette by site specific recombination, resulting in expression of marker 2. B.- Use of the one-way switch in a genetic screen for factor inducible genes. Cells with the switch vector inserted into their genome are transduced with a promoterless Cre gene trap vector. Insertions into active genes induce Cre expression which recombines the switch. Thus, at this point, selection for marker 1 eliminates all cells with gene trap insertions into expressed "housekeeping" genes leaving insertions into silent genes behind (1. OFF). Exposure to the stimulus of interest induces the expression of gene traps inserted in regulated genes, which activates the switch (2. ON). As a result, the cells express the downstream marker gene and can be selected in appropriate conditions. Since the switch recombination is irreversible, once activated, expression of marker 2 becomes independent of the trapped cellular promoter (3. ON or OFF), this enables the recovery of genes that are only transiently expressed.

The intrinsic properties of the gene traps make them complementary to RNA-based approaches such as microarray. First, trapping generates a single-cell reporter of the transcriptional activity rather than assessing mRNA abundance in the cell (Medico et al, 2001). Second, the high sensitivity of some gene traps like the one based on Cre allows the detection of transcribed genes even at extremely low levels. Third, Cre integration indicates the transcriptional direction of the gene where is integrated allowing identification of natural antisense transcripts.

1.5. Aim of the project

This thesis project is focused on a new approach to identify and characterize genes involved in the cell survival and hence in the process of conversion from normal cells to malignancy. The strategy is based on the identification of genes specifically induced by TNFα in a breast cancer cell line.

TNFα is not only a modulator of the immune response and thereby important in the inflammatory processes, but is also one of the strongest inducers of NF-κB dependent transcription. Indeed, TNFα induced apoptosis observed in some cells is preceded by transcriptional activation of survival genes. Identifying these transiently expressed genes is the aim of the present thesis; therefore a gene trap combined with Cre/*loxP* induced recombination was used for the identification of transiently activated genes.

Unraveling the signaling pathways in this cellular context could provide novel insights into pathways endowing a survival response. Alterations in these pathways could be related to carcinogenesis and therefore provide with potential new markers for diagnostic or targets for therapy.

2. - Materials

2.1. Machines and technical devices

Autoclave	VARIOKLAV® Dampfsterilisator, H+P Labortechnik GmbH, Oberschleiβheim, Germany
Bacterial incubator	B6030, Heraeus, Kendro Laboratory Products GmbH, Langenselbold, Germany
Bacterial shaker	Certomat®, HB. Braun, Melsungen, Germany
Centrifuges	Centrifuge 5415C, Eppendorf, Hamburg, Germany
	Megafuge 1.0R, Heraeus, Kendro Laboratory Products GmbH, Langenselbold, Germany
	Biofuge A, Heraeus, Kendro Laboratory Products GmbH, Langenselbold, Germany
	Sorvall® RC-5B, Du pont de Nemours GmbH, Bad Homburg, Germany
	Rotina 35 Hettich, Switzerland
Electrophoresis transfer cell	X CELLII™ Blot module, Invitrogen GmbH, Karlsruhe, Germany
Fluorescence microscope	Axioplan 2 imaging, Carl-Zeiss AG, Göttingen, Germany
Hybridization oven	Hybridizer HB-1000 UVP laboratory products, Cambridge, UK
	Hybridiser HB-ID, Techne, Burlington, USA
Incubator	BB6220, Heraeus, Kendro Laboratory products GmbH, Langenselbold, Germany
Laminar flow	DLF/REC4KL2A, Clean Air, Minneapolis, USA
Microplate reader	SPECTRA FLUOR PLUS, TECAN, Crailsheim, Germany
Microscope	DMIL Leica, Solms, Germany
Microscope camera	Axio Cam HRc, Carl-Zeiss AG, Göttingen, Germany
pH-meter	pH 210, microprocessor pH Meter, Hanna instruments, Kehl am Rhein, Germany
Phospho(r)imager	Molecular Dynamics Phospho(r)imager SI, Amersham Pharmacia Biotech, Freiburg, Germany
Power supply	E455 and E312 CONSORT, Tumhout, Belgium
	200/500, Carl-Roth GmbH+Co KG, Karlsruhe, Germany
RNA reader	Agilent 2100 Bionalyzer, Agilent technologies Inc, Böblingen, Germany
Scintillation counter	1500 TRI-CARB® Liquid scintillation analyzer Packard, Perkin Elmer LAS, Roggau-Jügesheim, Germany
Shaker	DUOMAX 1000, Heidolph instruments GmbH&Co KG, Karlsruhe, Germany
Spectrophotometer	Ultrospec III Pharmacia LKB, Freiburg, Germany
	Gene Quant II Pharmacia Biotech, Freiburg, Germany
Handcounter, contamination monitor	LB 122 Berthold, Bad Wildbad, Germany
Thermocyclers	MyCycler, Bio-Rad, Munich, Germany
	Gene amp PCR system 2400, Perkin Elmer, Wellesley, USA

	Robocycler gradient 96, Stratagene, Amsterdam, NL
Thermomixer	Comfort, Eppendorf, Hamburg, Germany
U.V. transiluminator	E.A.S.Y. RH-3, Herolab, Wiesloch, Germany
Vertical electrophoresis chambers	Ei9001 X-CELLII Mini Cell, Invitrogen GmbH, Karlsruhe, Germany
Vortex-mixer	72020 neolab, Heidelberg, Germany
Water bath	Shaking water bath 1083 GFL, Eppelheim, Germany
X-Ray film developing machine	KODAK M35 X-OMAT Processor, KODAK, Stuttgart, Germany

2.2. General laboratory supplies

Cromatography paper	3MM Chr, Whatman, Kent, UK
Filters	Millipore, Eschborn, Germany
Membranes	Trans-blot transfer Medium pure nitrocellulose membrane 0,2µM, BIO-RAD, Munich, Germany
	Nitrocellulose Hybond N+, Amersham, Freiburg, Germany
X-ray films	BioMax MS Film, KODAK, Sigma-aldrich, Taufkirchen, Germany
	CRONEX 5, AGFA, Agfa-Gevaert, Cologne, Germany
Plastic cuvettes	Sarstedt, Nümbrecht, Germany
cell culture plates	Costar GmbH, Bodenheim, Germany
	Greiner-bio-one GmbH, Frickenhausen, Germany
Pre-cast SDS-PAGE gels	NuPAGE™10% Bis-Tris Gel 1,0mmx15 gel, Invitrogen GmbH, Kalrsruhe, Germany

2.3. Kits

Cell proliferation kit II XTT	Roche, Mannheim, Germany
Dynabeads® mRNA purification kit	DYNAL biotech, Hamburg, Germany
DNeasy® Tissue kit	Qiagen, Hilden, Germany
RNeasy ® mini kit	Qiagen, Hilden, Germany
QIAquick® PCR purification kit	Qiagen, Hilden, Germany
QIAquick® Gel extraction kit	Qiagen, Hilden, Germany
Rediprime II™ random prime labeling system	Amersham, Freiburg, Germany
RNA 6000 Nanolab Chip® with reagents	Agilent technologies, Böblingen, Germany
Gen elute™ high performance (HP) plasmid midiprep kit	Sigma-Aldrich, Taufkirchen, Germany

2.4. Chemicals

All chemicals used were purchased from Calbiochem, Applichem, Merck, Boehringer-Mannheim, Becton Dickinson and Sigma-Aldrich Chemie GmbH.
MiliQ18MΩ or HPLC water was used in all procedures if required.

2.5. Enzymes

2.5.1. Restriction enzymes

Restriction enzymes were purchased from Invitrogen GmbH, Karlsruhe and New England Biolabs, Frankfurt am Main, Germany.

2.5.2. Modifying enzymes

DNA polymerases:	Taq DNA polymerase Recombinant Platinum® Taq DNA polymerase Platinum® Taq DNA polymerase High fidelity	DNA polymerases, reverse transcriptase, ribonulease H and ligase were from Invitrogen GmbH, Kalrsruhe, Germany
Reverse transcriptase:	SuperScript™II Reverse transcriptase	
Ligase:	T4 DNA ligase	
Nucleases:	Ribonuclease H	
	RQ1 Rnase-free Dnase	Promega, Mannheim, Germany
Other:	Calf intestinal phosphatase (CIP) Antarctic phosphatase Terminal deoxynucleotidyl-transferase T4 polynucleotide kinase	New England Biolabs, Frankfurt am Main, Germany

2.5.3. Enzyme inhibitors

RNAse inhibitors:	RNAsin	Promega, Mannheim, Germany
Protease inhibitors:	Aprotinin	Sigma-aldrich, Taufkirchen, Germany
	PMSF	Applichem, Darmstadt, Germany
	Leupeptin	Sigma-aldrich, Taufkirchen, Germany
	NaF	Sigma-aldrich, Taufkirchen, Germany
	Na_3VO_4	Sigma-aldrich, Taufkirchen, Germany

2.6. Pre-mixed solutions

10xtrypsin EDTA	Gibco, Invitrogen GmbH, Kalrsruhe, Germany
6xloading dye solution (for RNA)	Fermentas, St. Leon-Rot, Germany
Acrylamid solution (30%), mix 37.5:1 acryalmid: bis acrylamid	Applichem, Darmstadt, Germany
L-glutamine 200 mM	Gibco, Invitrogen GmbH, Kalrsruhe, Germany
NU PAGE® Antioxidant,	Invitrogen GmbH, Kalrsruhe, Germany
NU PAGE® LDS sample buffer	Invitrogen GmbH, Kalrsruhe, Germany
NU PAGE® MOPS SDS Running Buffer 20x	Invitrogen GmbH, Kalrsruhe, Germany
Ponceau's concentrate 2%	Sigma-aldrich, Taufkirchen, Germany
Protein Assay Reagent	Bio-Rad, Munich, Germany
Restore™ Western blot stripping buffer	Pierce, Bonn, Germany
Sodium-pyruvate MEM 100 mM	Gibco, Invitrogen GmbH, Kalrsruhe, Germany
Trypan blue stain 0.4%	Gibco, Invitrogen GmbH, Kalrsruhe, Germany

2.7. Reagents

FuGENE 6 transfection reagent	Roche, Mannheim, Germany
ULTRAhyb™	Ambion, Cambridgeshire, UK
TransPass D1 DNA trasfection reagent	New England Biolabs GmbH, Frankfurt am Main, Germany

2.8. Cell culture medium

RPMI 1640 and DEMEM with/out phenol red were purchased from Gibco, Invitrogen GmbH, Kalrsruhe, Germany

2.9. Cytokines

Recombinant human TNFα	R&D systems, Wiesbaden-Nordenstadt, Germany

2.10. Antibiotics

Antibiotic	Stock solution	Working concentration
Ampicillin sodium salt	100 mg/ml in water	100 µg/ml
Kanamicin	50 mg/ml in water	50µg/ml
Blasticidin S	5 mg/ml in water	10 µg/ml for MCF-7 cells
Geneticin sulphate (G418)	100 mg/ml in PBS	1 mg/ml for MCF-7 cells
Penicillin-streptomicyn	10.000 units/ml penicillin G sodium and 10.000 µg/ml streptomycin sulphate in 0.85% saline	100 units/ml penincillin G 100 µg/ml streptomycin
Puromycin	1 mg/ml in PBS	1 µg/ml in MCF-7 cells

2.11. Molecular weight standards

DNA Ladder:	1 kb Plus DNA Ladder,	All Ladders were from Invitrogen GmbH, Kalrsruhe, Germany
RNA Ladder:	0.24 - 9.5 kb RNA Ladder	
Protein Ladder:	MultiMark® multicolored standard See Blue® Plus 2	

2.12. Desoxyoligonucleotides

Synthetic desoxyoligonucleotides were purchased from Carl-Roth GmbH+Co KG, Karlsruhe or Invitrogen GmbH, Karlsruhe; the sequences are listed in the annex.
Random primers (3 µg/µl) were from Invitrogen GmbH, Kalrsruhe, Germany

2.13. Nucleotides

dNTPs 100 mM, Invitrogen GmbH, Kalrsruhe, Germany
[α-^{32}P]dCTP 3000Ci/mmol, 10mCi/ml, Hartmann Analytic GmbH, Braunschweig, Germany
ATP 10 mM, Invitrogen GmbH, Kalrsruhe, Germany

2.14. Antibodies

Antibodies were purchased from Dianova, Hamburg; Santa Cruz, Heidelberg; MoBiTec, Göttingen and Invitrogen, Karlsruhe.

2.15. Vectors

The plasmids used in this work have as selection marker genes conferring resistance against the antibiotics ampicillin (*bla* gene) or kanamycin (*nptII* gene from transposon Tn5).

2.15.1. Commercial vectors

pcDNA™6/V5-HisA, Invitrogen: plasmid for the expression of fusion proteins with a V5 epitope and a 6xHis tag in eukaryotes.

pVPack-GP vector, Stratagene: helper plasmid required for the production of Mouse Moloney Leukemia virus (MMLV) based retroviral particles. Expression vector for *gag* and *pol* genes coding for group specific antigen and reverse transcriptase/integrase.

pMD2.G: helper plasmid required for the production of Mouse Moloney Leukemia virus (MMLV) based retroviral particles. Expression vector for the VSV-G envelope, http://tronolab.epfl.ch/page58115.html.

RZPD clone: IRAUp969D0665D/-IMAGE clone. This plasmid contains a *ZFP67* cDNA with the complete coding region.

RZPD clone: IRAUp969F0776D/-IMAGE clone. This plasmid contains a *FLJ14451* cDNA with the complete coding region.

2.15.2 Provided vectors

pBabeSIN: Mouse Moloney Leukemia virus (MMLV) based retroviral vector in which all transcriptional regulatory sequences from the U3-region have been removed.

pBabeU3CreΔBII(-) provided by Joachim Altschmied: this plasmid is used for the production of U3Cre gene trap retrovirus particles; the *Cre* gene is located in the retroviral U3 region.

pcDNA6egfpIRESblas provided by Joachim Altschmied: pcDNA™6/V5-HisA with a SV40 driven a blasticidin IRESegfp cassette allowing selection and fluorescent visualization of transfected cells.

SP100gfp provided by Hans Will, University of Hamburg: expression vector for a SP100gfp fusion protein.

PML-1gfp provided by Hans Will, University of Hamburg: expression vector for a PML-1gfp fusion protein.

2.15.3. Constructed vectors

pneoPGKlxpurolxdnFADD: the plasmid contains a PGK driven floxed puro-pA cassette and silent dnFADD-pA cassette; additionally it contains a SV40 promoter driven *neo* selection marker.

pneoPGKlxdnFADD: the plasmid is a deletion mutant of pneoPGKlxpurolxdnFADD, in which the lxpuro-pA cassette was removed by *in vitro* recombination with Cre recombinase.

pBSCMVlxtkneolxdnFADD: this plasmid carries the one way gene expression switch, it has a CMV promoter driven floxed tkneo-pA gene cassette and a silent dnFADD-pA cassette.

vCMVlxtkneolxdnFADD: plasmid for the production of retroviral particles containing the CMVlxtkneolxdnFADD switch reporter cassette inserted in antisense orientation between the LTRs of a pBabeSIN vector.

pBabepuroPGKcreS: retroviral *Cre* expression vector with PGK driven *Cre* and SV40 driven *puromycin-acetyl-transferase* selection marker.

pCMVZFP67egfpIRESblas: eukaryotic expression vector for a V5-His fusion protein and SV40 driven *blasticidin deaminase* selection marker and egfp.

pCMVZFP67neo: eukaryotic expression vector for a V5-HisA fusion protein and SV40 driven *neomycin-phosphotransferase* selection marker.

pCMVFLJ14451egfpIRESblas: eukaryotic expression vector for a V5-HisA fusion protein SV40 driven *blasticidin-deaminase* selection marker and egfp.

pCMVFLJ14451neo: eukaryotic expression vector for a V5-HisA fusion protein and SV40 driven *neomycin-phosphotransferase* selection marker.

2.16. Software

General:	Photoshop 4.0, Adobe Systems Inc. Microsoft Office 2001 for Macintosh
DNA analysis:	ABI Prism DNA Sequencing Software, Perkin Elmer ABI DNA Strider™1.2, CEA
PCR primer prediction:	Oligo® 4.04-s, National Biosciences, Inc Primer3, http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi
RNA analysis:	2100 Expert software
Densitometer software:	Image Quant, Molecular Dynamics Quantity one, Bio-Rad, Munich

2.16.1. On line bioinformatics resources

Protein motifs, patterns, profiles analysis:

ScanProsite http://www.expasy.ch/tools/scnpsite.html

Database similarity search software:

BLAST http://www.ncbi.nlm.nih.gov/BLAST

DNA bioinformatics data bases:

Gen Bank http://www.ncbi.nlm.nih.gov/Genbank/GenbankSearch.html
ENSEMBL http://www.ensembl.org
Celera http://www.celera.com

Source	http://genome-www5.stanford.edu/cgi-bin/source/sourceSearch
Harvester	http://harvester.embl.de

Microarray database

Oncomine	http://www.oncomine.org:80/main/mainx.jsp

Promoter and first exon prediction routines

Dragon promoter finder version 1.5	http://research.i2r.a-star.edu.sg/promoter/promoter1_5/DPF.htm
Neural Network Promoter Prediction	http://www.fruitfly.org/seq_tools/promoter.html
FirstEF	http://rulai.cshl.org/tools/FirstEF/
McPromoter MM:II	http://genes.mit.edu/McPromoter.html
promoter 2	http://www.cbs.dtu.dk/services/Promoter/
promoter scan	http://thr.cit.nih.gov/molbio/proscan/

3. - Methods

3.1. Purification, manipulation and detection of DNA

3.1.1. DNA storage

DNA was stored in TE buffer at –20°C.

TE-buffer:

10	mM	Tris•Cl, pH 7.6
1	mM	EDTA

Autoclave, store at room temperature

3.1.2. Plasmid DNA isolation

Plasmid DNA was purified with the QIAprep® Spin Miniprep kit (Qiagen, Hilden) starting from 1.5-2 ml dense overnight culture, if less than 10 µg were required or, for larger amounts, from 100 ml culture using the Gen elute™ high performance (HP) plasmid midiprep kit (Sigma-aldrich, Taufkirchen). Purification was performed as described by the suppliers.

3.1.3. DNA isolation from eukaryotic cell lines

DNA isolation was performed with the DNAeasy Tissue kit (Qiagen, Hilden) according to manufacturer's specifications.

3.1.4. Measurement of DNA concentration

DNA concentration and purity was determined spectrophotometrically. Absorption of 1 at 260 nm corresponds to a concentration of 50 µg/ml double strand DNA, the ratio between the absorptions at 260 nm and 280 nm for DNA should be between 1.7 and 2.0 (Maniatis et al., 1982).

3.1.5. Enzymatic manipulation of DNA

Enzymatic modifications of DNA were done with commercial enzymes using buffer and incubation conditions suggested by the manufacturers.

3.1.5.1. Endonucleolytic digestion

Different restrictions enzymes were used, incubation time was dependent on the amount of DNA and enzyme concentration, as a rule of thumb 1 U enzyme should digest 1 µg DNA within an hour. Incubation temperatures were chosen as specified for each enzyme.

3.1.5.2. Dephosphorylation

5' terminal phosphate groups were removed from DNA with calf-intestinal-phosphatase or antarctic phosphatase according to the manufacturer's specifications.

3.1.5.3. Ligation

T4 DNA ligase was used for the ligation of DNA fragments. For cloning purposes, approximately 100-200 ng DNA with a vector:insert ratio of 1:1 and 1 U ligase were combined in a final volume of 20 µl ligation buffer and incubated overnight at 16°C.

3.1.5.4. Filling in of double stranded DNA with 5'-overhangs

The enzyme used for this purpose was the Klenow fragment of *E. coli* DNA polymerase I.

3.1.6. Electrophoresis of DNA in agarose gels

Double stranded DNA fragments can be separated according to their lengths by gel electrophoresis through agarose gels, with the separation range being dependent on the agarose concentration (0.6-2% w/v). Agarose was dissolved in 1xTAE buffer by boiling in a microwave oven. When the molten gel was cooled to 50°C ethidium bromide was added to a

final concentration of 0.1 µg/ml, thoroughly mixed by gentle swirling and then poured into an electrophoresis tray with an appropriate comb. After solidification the tray was transferred to a horizontal electrophoresis chamber and the gel covered with 1xTAE buffer. DNA samples were mixed with loading buffer and loaded into the gel. The gels were run at 5-8 V/cm, after gel electrophoresis the DNA was visualized on an U.V. transilluminator.

TAE buffer:

40	mM	Tris•Cl pH 8.5
40	mM	glacial acetic acid
2	mM	EDTA

6xloading buffer:

30	%	glycerol
10	mM	EDTA
0.1	%	SDS
0.25	%	xylene cyanol FF
0.25	%	bromophenol blue

3.1.7. DNA purification

DNA was purified from aqueous solutions with the QIAquick® PCR purification kit and recovered from agarose gels using the QIAquick® gel extraction kit (both Qiagen, Hilden) following the manufacturer's instructions.

3.1.8. Polymerase chain reaction (PCR)

A mix was prepared with 0.2 mM dNTPs, 1.5 mM $MgCl_2$, 1.5 units heat stable DNA polymerase (Taq polymerase from *Thermus aquaticus*), 10 pmol each primer and DNA template to 30-50 µl end volume in a PCR tube. Amplification was performed in a thermocycler where an initial denaturation of 5 min was followed by 30 to 40 cycles of denaturation-annealing-extension steps. Denaturation and extension steps were performed for 30 sec, a rule of thumb for the extension time was 1 min for 1 kb. Denaturation temperature was 95°C, extension temperature 72°C for Taq DNA polymerase and 68°C for proofreading polymerases. Annealing temperature was adjusted according to primer composition and was 4°C below the melting temperature of the primer template hybrid (Tm).

Primers were designed using commercial (Oligo) or publicly available (Primer 3) software. They were between 17-28 nucleotides long, with one G or C at the 3'-end and Tm between 55°C and 65°C. The melting temperature was calculated by the formula Tm=4x(G+C)+2x(A+T); where A is adenosine, T is thymidine, C is cytidine and G is guanosine.

The amount of template was dependent on the type of amplification. For a colony PCR one colony was picked into 20 µl LB medium and 5 µl were used as template in the PCR reaction. Amplifications with DNA templates were performed with 500 ng genomic DNA, 0.2 to 2 ng plasmid DNA or, in the case of reverse transcriptase PCRs (RT-PCRs), with 2 to 5% of a cDNA synthesis reaction starting from 1 µg total RNA.

3.1.9. Inverse PCR

Genomic DNA (2 to 4 µg) was digested overnight with 10 units *Pst*I or *Nsp*I. Restriction enzymes were inactivated for 30 min at 80°C, DNA fragments were purified with the QIAquick® PCR purification kit (Qiagen, Hilden) and eluted in 40 µl HPLC-water. DNA fragments were ligated overnight at 16°C with 3200 units T4 DNA ligase (Invitrogen GmbH, Karlsruhe) in a 100 µl ligation reaction. Ligase was inactivated for 15 min at 65°C, the reaction products purified with the QIAquick® PCR purification kit (Qiagen, Hilden) and eluted in 30 µl HPLC-water. A first PCR was performed under standard conditions with 10 µl ligation products in 50 µl volume with the addition of DMSO to 2% (v/v) and 2.5 U Platinum Taq polymerase (Invitrogen GmbH, Karlsruhe). After that, a nested PCR was performed under identical conditions with 0.1 to 1 µl from the first PCR as template. The reaction products were analyzed in an agarose gel and sequenced after purification with the PCR purification kit (Qiagen, Hilden) if only one product was present (Hui, Wang et al. 1998). The following primers were used:

First PCR	Cre43/SY2	*Pst*I digested DNA
	CreiPCR rev2/iPCRU3	*Nsp*I digested DNA
Second PCR	Cre1P/SY1	*Pst*I digested DNA
	CreiPCRrev1/iPCRU4	*Nsp*I digested DNA
Sequencing	Cre 1P	*Pst*I digested DNA
	Cre 43	*Nsp*I digested DNA

3.1.10. Radioactive labeling of DNA fragments

Probes were either labeled by random primed labeling or asymmetric PCR. For random primed labeling around 500 ng of a DNA fragment were radioactively labeled with 5 µl (0.33 µM end concentration) [α-^{32}P]-dCTP (3000Ci/mmol-10mCi/ml, Hartmann Analytic GmbH Braunschweig), in a final volume of 50 µl using the Rediprime II™ random prime labeling system (Amersham, Freiburg) following manufacturer's recommendations.

Single strand radioactive probes were synthesized by linear amplification in an asymmetric PCR, in which only one primer was used and the unlabelled dCTP was replaced by a mixture of cold and radioactive dCTP, to end concentrations of 1.2 µM cold dCTP and 0.33 µM [α-^{32}P]-dCTP (3000Ci/mmol-10mCi/ml, Hartmann Analytic GmbH Braunschweig). As template 10-100 ng of the DNA fragment to be labeled were used, amplification was carried out 40 cycles with an extension time of 4 min.

Non incorporated nucleotides were removed by chromatography using Micro spin S-200 HR columns (Amersham, Freiburg) following manufacturers instructions. Activity of the radioactive probes was measured with 1 µl radioactive-labeled DNA in a Scintillation counter.

3.2. Purification, manipulation and detection of RNA

For RNA work a series of precaution were taken to avoid RNase contamination. RNA was stored at –80°C, always transported on ice, gloves were always used, pipette tips were autoclaved or RNase free pipette tips used, glassware was autoclaved and water was treated with DEPC.

DEPC water:

0.1	%	DEPC

Stir overnight, autoclave, store at room temperature

3.2.2. RNA isolation from eukaryotic cell lines

Cells were grown in 9 cm cell culture dishes until they reached 80-90% confluence, trypsinized as described and collected by centrifugation (see section 4.2). Cells were then

resuspended, centrifuged at 210 g for 5 min and rinsed once with PBS. The cell pellet could be stored at –20°C up to 2 weeks before the RNA was isolated.

RNA from cell pellets was extracted using the RNeasy mini kit (Qiagen, Hilden), which can be used to isolate molecules longer than 200 nucleotides, following manufacturer's recommendations.

3.2.3. Measurement of RNA concentration

RNA concentration was determined in an Agilent 2100 Bionalyzer (Agilent technologies Inc, Böblingen) with a RNA 6000 Nanolab Chip® (Agilent technologies Inc, Böblingen) following the manufacturer's recommendations.

3.2.4. RNA concentration

RNA can be concentrated with a vacuum centrifuge ("speed-vac"), which evaporates water from the sample. If the RNA volume to be concentrated was too large, RNA was precipitated with ammonium acetate. The ammonium acetate concentration was adjusted to 0.5 M with a 5 M stock solution, then 2.5 volumes ethanol were added and the sample was incubated at –80°C for at least 30 min before centrifugation for 30 min in a tabletop centrifuge at full speed and 4°C. The precipitate was resuspended in DEPC water.

3.2.5. mRNA isolation

mRNA was isolated from total RNA using the Dynabeads® mRNA purification kit (Dynal biotech, Hamburg) following manufacturers instructions.

3.2.6. DNA digestion

DNA was removed from the RNA sample with RQ1 RNase-free DNase (Promega, Mannheim) following the manufacturer's instructions. Briefly, 2 µg RNA were incubated with 2 units RQ1 RNase-free DNase in a 10 µl end volume reaction for 30 min at 37°C. To terminate the digestion 1 µl RNase stop solution was added and the reaction was incubated for 10 minutes at 65°C.

3.2.7. First strand cDNA synthesis

RNA was reverse transcribed using Superscript II reverse transcriptase (Invitrogen GmbH, Kalrsruhe) and random hexamer primers (Invitrogen GmbH, Kalrsruhe) following manufacturers instructions. Briefly, 2 µg RNA were incubated for 5 min at 65°C with 500 ng random hexamers, 20 pmol dNTPs and DMSO (0-9%) in a final volume of 24 µl. The sample was then placed on ice, 2 µl RNasin 40 U/µl (Promega, Mannheim), 4 µl DTT 0.1 M and first strand buffer were added incubated for 10 min at 25°C and afterwards for 2 min at 42°C. 2 µl Superscript II (400 U) were added to the reaction and incubation continued for 90 min at 42°C. cDNA synthesis was terminated with a 15 min incubation step at 70°C, afterwards the RNA was removed by incubation with 2 units RNase H for 20 min at 37°C.

3.2.8. Northern blot

3.2.8.1. RNA electrophoresis

3.2.8.1.1. Gel preparation

RNAs can be separated according to their length on a denaturing gel. For a 150 ml gel, 1.5 g agarose were dissolved in water by boiling in a microwave oven and cooled to 55°C. 15ml 20xMOPS and 9 ml formaldehyde 37% were added in a chemical fume hood thoroughly mixed by gentle swirling and then poured into an electrophoresis tray with an appropriate comb. After solidification the tray was transferred to a horizontal electrophoresis chamber and the gel covered with 1xMOPS electrophoresis buffer.

3.2.8.1.2. Sample preparation

RNA samples were processed in parallel to a RNA ladder, the maximum sample volume was 6 µl, if more dilute, samples were concentrated by precipitation. A maximum of 30 µg RNA per lane was used (lane width 0.6 cm); more RNA is not recommended as the gel is overloaded with the consequence of a bad RNA separation. The 6 µl RNA was mixed with 15 µl denaturing buffer, incubated at 65°C for 20 min and chilled on ice. 4 µl 6xRNA loading dye solution (Fermentas, St. Leon-Rot) were added to the sample and 22 µl of the mixture were loaded on the agarose-formaldehyde gel. Gels were run at 6.6 V/cm for 3 hours.

20xMOPS electrophoresis buffer:

400	mM	MOPS
100	mM	Na-acetate
10	mM	EDTA

Autoclave, store at room temperature

Denaturing buffer:

2	x	MOPS electrophoresis buffer
50	%	formamide (v/v)
6.5	%	Formaldehyde (v/v)

Always make freshly

3.2.8.2. Transfer to nitrocellulose membranes

Electrophoretically separated RNA was transferred to a Nitrocellulose Hybond N+ (Amersham, Freiburg) membrane by capillary force using 10xSSC as transfer buffer. The gel was equilibrated in 10xSSC for 15 min and placed head over on a Whatman paper bridge, which contacts a reservoir containing 10xSSC. A membrane of the same size as the gel was placed on top of the gel, followed by 2-4 layers of Whatman also cut to gel size, a pile of paper and a 0.5 kg weight on top. The transfer to the nitrocellulose membrane was done overnight.

20xSSC:

3	M	NaCl
300	mM	Na citrate

Autoclave, store at room temperature

3.2.8.3. Staining of RNA on nitrocellulose membranes

After the transfer, the membrane was stained with methylene blue to check the transfer efficiency. Therefore, the membrane was incubated with the staining solution for 5 min at room temperature and washed with an excess of water. When using total RNA, two bands appear corresponding to the highly abundant 28S and 18S ribosomal RNAs. If mRNA was transferred to the membrane only weak bands with the residual 28S and 18S can be visualized. The methylene blue was washed out with washing solution 2.

Staining solution:

0.04	%	methylene blue (w/v)
500	mM	Na-acetate

Washing solution 2:

0.1	x	SSC
0.1	%	SDS (w/v)

3.2.8.4. Hybridization

3.2.8.4.1. Hybridization probe

Probes to detect abundant RNAs like β-actin mRNA were radioactively labeled by random priming (section 3.2.10), probes for less abundant RNAs by asymmetric PCR (section 3.2.10).

For normalization all Northern blots were hybridized with a β-actin probe (Thellin, Zorzi et al. 1999).

3.2.8.4.2. Hybridization to the membrane

After staining with methylene blue and distaining, the membrane was prehybridized with hybridization buffer for 1 hour at 42°C. Then, the denatured probe was added (10^6-$2x10^6$cpm) and hybridized overnight at 42°C or two hours for abundant RNAs like actin mRNA or 18S rRNA. To remove unbound probe the filter was washed twice for 5 min with washing solution 1 at 42°C and once for 5 min with washing solution 2 at 42°C. A final washing step with washing solution 2 for 30 min at 60°C ensured removal of all unspecifically hybridized probe.

Hybridization buffer:

ULTRAhyb™ Ambion, Cambridgeshire

Washing solution 1:

2	x	SSC
0.1	%	SDS (w/v)

Washing solution 2:

0.1	x	SSC
0.1	%	SDS (w/v)

3.2.8.5. Detection

After washing, the membrane was sealed in a transparent plastic bag with 2 ml washing solution 2. The plastic bag was placed in an X-ray cassette and a BioMax MS Film, KODAK (Sigma-aldrich, Taufkirchen) was exposed to the membrane. Exposition time was dependent on signal strength. In general an overnight exposition was done and further expositions if necessary; for quantifying, the signals were recorded with a Phospho(r)imager.

3.2.8.6. Stripping

For further hybridizations, the membrane was immersed in boiling 0.1% SDS solution and cooled to room temperature in this solution.

3.3. Preparation, storage and detection of proteins

3.3.2. Preparation of cell lysates/extracts

Cells from a 9 cm plate at 70-80% confluence were detached from the growth surface with trypsin as described in section 3.6.2. Cells were centrifuged at 210 g for 5 min and rinsed once with PBS. The cells were resuspended in 150-200 µl lysis buffer and disrupted mechanically by vortexing. Proteins were recovered from the supernatant after centrifugation at full speed in a tabletop centrifuge for 10 min and stored at –20°C.

Lysis buffer:

50	mM	Tris•Cl, pH 7.4
150	mM	NaCl
0.1	%	NP-40 (v/v)
10	%	glycerol (v/v)
25	mM	Na-β-glycerophosphate

Store at 4°C

Directly before use proteinase inhibitors were added from concentrated stock to the following final concentrations:

1	mM	PMSF, stock solution 100 mM in ethanol
10	mM	Na_3VO_4, stock solution 2 M in water

5	µM	Leupeptin, stock solution 5 mM in water
20	mM	NaF, stock solution 1M in water
8	µg/ml	Aprotinin, stock solution 1.6 mg/ml in 0.9% NaCl and 0.9% benzyl alcohol

3.3.3. Measurement of protein concentration

Protein concentration in lysates was determined with the Bio-Rad protein assay (Bio-Rad, Munich) following manufacturers recommendations. Briefly, 1 ml Bio-Rad protein assay solution (diluted 1:5 with water and filtered through a paper filter) was mixed with sample (3-18 µg) and the absorbance measured at 595 nm. Sample concentration was calculated from a standard curve obtained with BSA (Bradford 1976).

3.3.4. Western blot

3.3.4.1. Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)

3.3.4.1.1. Gel preparation

Proteins can be separated largely on the basis of their relative molecular mass by electrophoresis in polyacrylamid gels under denaturing conditions. SDS-polyacrylamid gels consisted of a separation gel and a stacking gel on top of it. After all components of the separation gel were mixed, ammonium persulfate and TEMED were added to initiate polymerization. The polymerization mix was poured in a vertical electrophoresis cassette and overlayed carefully with water. After complete polymerization of the separation gel, the water was poured off. The 5% stacking gel was poured on top and a comb was inserted to form the sample slots.

Separation gel:

6-15	%	Acrylamid: bisacrylamid (v/v)
0.1	%	SDS (w/v)
400	mM	Tris•Cl pH 8.8
0.1	%	Ammonium persulfate (w/v)
0.08	%	TEMED (v/v)

Stacking gel:

5	%	Acrylamid: bisacrylamid (v/v)
0.1	%	SDS (w/v)
130	mM	Tris•Cl pH 6.8
0.1	%	Ammonium persulfate (w/v)
0.1	%	TEMED (v/v)

3.3.4.1.2. Electrophoresis

Protein samples (20-50 µg) were mixed with loading buffer at a ratio of 2.5:1, boiled for 10 min and chilled on ice. Samples were briefly centrifuged and loaded onto the gel. For molecular weight determination, a protein molecular weight marker was loaded parallel to the samples. The gel was run for 2-3 hours at 100-120 V constant voltage, until the bromophenol blue had nearly reached the bottom of the gel.

Loading buffer:

Add 5 % mercaptoethanol (v/v) to the loading buffer (NuPage® LDS sample buffer, Invitrogen) before use, can be stored at –20°C

Running buffer:

25	mM	Tris base
250	mM	Glycin
0.1	%	SDS (w/v)

Store at room temperature

3.3.4.2. Electroblotting

Electrophoretically separated proteins were blotted onto nitrocellulose membranes (Trans-blot transfer Medium pure nitrocellulose membrane 0.2 µM, BIO-RAD, Munich) in Towbin transfer buffer using a NOVEX Western apparatus following manufacturer's instructions. Briefly, after disassembling the gel electrophoresis cassette, the stacking gel was cut off and the separation gel was assembled into a sandwich together with the nitrocellulose membrane. This sandwich was prepared in a tank with transfer buffer containing the following components, all prewetted with transfer buffer, in the given order: 2 blots pads, 1 piece of Whatman paper, gel, 0.2 µM pure nitrocellulose membrane (BIORAD, Munich), 1 piece of

Whatman paper and 3-4 blot pads; all papers and filters were cut to the dimension of the gel. The gel/membrane assembly was held securely between the two halves of the blot module ensuring complete contact of all components. The blot module was filled with transfer buffer and run at 35 V for 90 min.

Towbin transfer buffer:

12	mM	Tris base
96	mM	glycine
20	%	methanol (v/v)

3.3.4.3. Staining

After blotting was completed, transfer and equal loading of the proteins were controlled by staining the membrane with 0.1% Ponceau red (Sigma-aldrich, Taufkirchen) 3 min at room temperature. The membrane was distained by washing 10-15 min with TBST before the immunostaining was performed.

3.3.4.4. Immunological detection

Proteins can be detected on the membrane with antibodies, which bind to a specific region on the protein (epitope). All antibodies used were diluted according to manufacturer's instructions in TBST-5% low-fat milk powder (w/v). The incubations and washes were done at room temperature on a shaker, if not indicated otherwise.

The membrane was blocked in TBST-5% low-fat milk powder (w/v) for at least 1 hour at room temperature or overnight at 4°C. After blocking the membrane was placed bubble free in a plastic bag with 5 ml primary antibody and incubated for 1 hour. This incubation was followed by three washes with TBST (5 min each) to remove unbound antibody. After incubation with the secondary antibody conjugated to horseradish peroxidase for 1 hour the membrane was washed as before. The membrane was rinsed with water to replace TBST and incubated in the dark for 1 min with developing solution, which was made freshly in the dark by mixing ECL solutions 1 and 2 (1:1). Afterwards, the membrane was placed in a transparent plastic bag and exposed to an X-ray. Oxidation of luminol by horseradish peroxidase in the presence of hydrogen peroxide leads to emission of photons, which can be detected by a light

sensitive film. Depending on signal strength the exposure ranged from 5 sec to 30 min, after which no further light is emitted.

After protein detection the membrane was stored wet at 4°C and could be reused again. For the detection of other proteins/antigens the antibody was stripped off in stripping solution (Restore^{TM} Western blot stripping buffer, Pierce, Bonn) at room temperature for 20 min. The stripping solution was discarded and the membrane washed several times with TBST. The membrane was blocked again and the detection repeated with another antibody.

	Antibody	Commercial supplier	Working dilution
Primary	Anti FADD	MoBiTec, Göttingen	1:1000
	Anti V5	Invitrogen, Karlsruhe	1:5000
	Anti tubulin Ab-4 (DM1A+DM1B)	Dianova, Hamburg	1:1000
Secondary	Anti mouse	Dianova, Hamburg	1:2000
	Anti mouse	Santa Cruz, Heidelberg	1:1000
	Anti rabbit	Dianova, Hamburg	1:2000

All the antibodies were diluted in 5% low-fat milk

TBST buffer:

150	mM	NaCl
50	mM	Tris•Cl pH 8.0
0.05	%	Tween 20 (v/v)

ECL solution 1:

2.5	mM	luminol
400	µM	p-coumaric acid
100	mM	Tris•Cl pH 8.5

ECL solution 2:

0.02	%	H_2O_2 (v/v)
100	mM	Tris•Cl

Store ECL solution 1 and 2 at 4°C

Stock solutions:

250	mM	Luminol in DMSO
90	mM	p-coumaric acid in DMSO

Store at –20°C

3.4. Working with bacteria

3.4.1. Bacterial strains

The *E.coli* bacterial strains XL-1 blue, DH5-alpha and XL10 were used. All of them have a deletion in the gene for β-galactosidase, which can be compensated with a plasmid vector encoding the α-peptide fragment of the β-galactosidase (α-complementation; Langley, Villarejo et al. 1975). Such plasmids carry a short segment of *E.coli* DNA containing the regulatory sequences and the coding information for the first 146 aa of the *β-galactosidase* gene. The host encoded β-galactosidase and the plasmid encoded associate to form an enzymatically active protein and form blue colonies in the presence of X-Gal. Insertion of a foreign DNA fragment into the α-peptide coding region of the plasmid almost invariably results in the production of an amino-terminal fragment that is no longer capable of α-complementation and thus, bacteria harboring such a plasmid form white colonies in the presence of X-Gal.

The bacterial strains have the following genotypes (genes listed signify mutant alleles; genes on the F' episome, however, are wild-type unless indicated otherwise):

DH5-alpha	F- ϕ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F)U169 *rec*A1 *end*A1 *hsd*R17(r_k-, m_k+) *pho*A *sup*E44 *thi*-1 *gyr*A96 *rel*A1 λ-
XL10	Tet^r Δ(*mcrA*)*183* Δ(*mcrCB-hsdSMR-mrr*)*173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac*Hte [F´ *proAB lacI*^{q}ZΔ*M15* Tn*10* (Tet^r) Amy Cam^r]
XL1-blue	*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [F´ *proAB lacI*qZΔ*M15* Tn*10* (Tetr)].

3.4.2. Growth and storage

All bacterial strains, were grown at 37°C as shaking suspension cultures in LB medium or on LB agar plates, which were supplemented with antibiotics if required. Bacteria on agar plates were stored up to 4 weeks at 4°C. For long term storage at –80°C glycerol was added to LB-suspension cultures to a final concentration of 15% (v/v) (Ausubel et al 1989).

LB medium:

10	g/l	bacto tryptone
5	g/l	yeast extract
10	mM	NaCl

Autoclave, store at room temperature

Agar plates:

10	g/l	bacto tryptone
5	g/l	yeast extract
10	mM	NaCl
15	g/l	agar

Autoclave, cool to 55°C, add antibiotic if required and pour in Petri-plates, store in the dark at 4°C

3.4.3. Competent bacteria

Transformation of competent bacteria followed the protocol from Hanahan (Hanahan 1983), all the reagents were pre-cold at 4°C. A overnight 2 ml culture coming from a single colony of *E.coli* DH5α was used to inoculate a 50 ml culture which was grown overnight at 37°C at 200 rpm shaking. This culture was used next day for a 1500 ml culture that was grown at 37°C with shaking till the absorbance at 600 nm was between 0.4-0.6 O.D. The bacteria were then centrifuged at 1638 g at 4°C for 15 min. For all further steps pre-cold buffers were used and the manipulations and incubations were performed on ice. The bacteria were resuspended in 50 ml TFB-I buffer and incubated 90 min on ice, centrifuged as before, resuspended in 40 ml TFB-II buffer and flash frozen in liquid nitrogen in aliquots of 200 µl, frozen competent bacteria were stored at –80°C.

TFB-I:

15	%	glycerol (v/v)
30	mM	Calcium acetate
100	mM	RbCl
10	mM	$CaCl_2$
50	mM	$MnCl_2$

Adjust pH to 5.8 with acetic acid
2 M $CaCl_2$ and $MnCl_2$ stock solutions were prepared, autoclaved and used for this buffer

TFB-II:

15	%	glycerol (v/v)
10	mM	MOPS
10	mM	RbCl
75	mM	$CaCl_2$

Adjust pH to 7.0 with NaOH

3.4.4. Transformation with DNA

If commercially prepared competent bacteria were used, the manufacturer's instructions were followed. Competent bacteria were thawed on ice and 100 µl were mixed carefully with DNA (ligation reaction or plasmid DNA, up to 20 µl). The suspension was incubated for 30 min on ice, heat-shocked at 42°C for 90 sec before 1 ml pre-warmed LB medium was added and then incubated at 37°C for 20-40 min with vigorous shaking. Afterwards the suspension was centrifuged 4 minutes at 4000 rpm in a table centrifuge, resuspended in 60 µl LB medium and evenly spread on an agar plate. The agar plate was incubated 12-16 hours at 37°C.

Recombinant plasmids were identified by colony PCR or restriction digestion of purified DNA.

Plates:

Approximately 1 hour before use, spread 60 µl 10% X-Gal
and 50 µl 100 mM IPTG onto plate, air dray, not store

Stock solutions:

100	mM	IPTG stock solution
10	%	X-Gal

Store at –20°C

3.5. Work with mammalian cells

3.5.1. Cell lines

MCF-7 (ATCC-NR. HTB-22)

The MCF-7 cell line is a human cell line from a pleural effusion derived from a breast carcinoma; the cells have an epithelial morphology and grow adherent (Soule, Vazguez et al. 1973). Cultivation in various laboratories has given rise to different subpopulations (Prest, Rees et al. 1999; Devarajan, Chen et al. 2002) with distinct phenotypic features and responses to TNFα (Burow, Weldon et al. 1998).

HeLa (ATCC-NR. CCL-2)

The HeLa cell line is a human cervical-adenocarcinoma cell line. The cells have an epithelial morphology and grow adherent.

293 T

The 293T cell line is a human cell line from fetal kidney constitutively expressing the simian virus 40 (SV40) large T antigen. The cells have an epithelial morphology and grow adherent.

3.5.2. Routine culturing

All mammalian cell lines were grown at 37°C and 95% humidity in the presence of 5% CO_2. Cell lines were maintained at subconfluent densities in culture medium supplemented with 10% heat-inactivated fetal calf serum (BIOCHROM AG seromed®, Berlin), 100 units/ml penicillin, 100 µg/ml streptomycin, 1 mM sodium pyruvate and 2 mM glutamine (Invitrogen, Karlsruhe). Cell culture medium was change every 2-3 days and cells were passage at 80% confluence.

For passaging, cells were detached from the growth surface with 0.05% trypsin in PBS (Invitrogen, Karlsruhe). Briefly, medium was removed from the plate and the cells rinsed once with PBS. Trypsin solution was added, distributed evenly by tilting the plate and removed leaving only a thin film of the solution. The plate was then incubated at 37°C until cells detached. At this point medium containing serum was added to inhibit trypsin, cells were resuspended and splitted into new culture dishes. The split ratio was between 1:10 and 1:20 for all cell lines used.

The following basic media were used for the different cells lines and supplemented as described above:

Medium	Cell line
RPMI 1640 with L-glutamine	Hela
DMEM without phenol red, 4500 mg/l glucose without L-glutamine, without Na pyruvate	MCF-7
DMEM with phenol red, 4500 mg/l glucose with L-glutamine, without Na pyruvate	293-T

PBS buffer:

10	mM	$Na_3PO_4 \cdot 2H_2O$, pH7.4
1.7	mM	KH_2PO_4
137	mM	NaCl
2.7	mM	KCl

Autoclave, store at room temperature

3.5.3. Cell counting

For experiment requiring exact cell numbers, cells were counted using a Neubauer chamber. Therefore, cells were diluted 1/5 or 1/10 with trypan blue stain 0.4% (Gibco, Invitrogen GmbH, Kalrsruhe), which stains only dead cells, whereas living cells do not take up the stain (trypan blue exclusion).

3.5.4. Freezing and thawing of cells

For long term storage subconfluent cells were detached from the growth surface with trypsin as described. Then they were collected by centrifugation at 210 g for 5 min at room temperature and resuspended in freezing solution, (2 ml freezing solution for cells from a subconfluent 9 cm diameter dish). The cell suspension was transferred to appropriate cryotubes in 1 ml aliquots and these were placed into a cryocontainer with isopropanol and kept for 24 hours at –80°C. After that, the cells were transferred to liquid nitrogen for long term storage.

For thawing, cryotubes were placed in a 37°C water bath. Immediately after thawing, the cell suspension was transferred into a centrifuge tube containing 5 ml regular growth medium and centrifuged at 210 g for 5 min at room temperature. The cell pellet was resuspended in growth medium and transferred into a culture dish. The next day the medium was changed to remove dead cells.

Freezing solution:

10 % DMSO in fetal calf serum

3.5.5. Transfection of mammalian cells

Different transfection procedures and reagents were used depending on the transfected cell type; if stably selected transfectants were desired, selection for antibiotic resistance was started 2 days after transfection.

3.5.5.1. Transfection with FuGene 6 (Roche, Mannheim)

One day before the transfection $5x10^6$ MCF-7 or $4x10^6$ 293-T cells were seeded in a 9 cm cell culture dish. For transfection cells were washed and fresh supplemented medium was added. FuGene was prepared according to manufacturer's instructions. Briefly, for a 9 cm diameter cell culture dish, 12 µl FuGene were added to 400 µl medium without serum. After 5 minutes incubation at room temperature, 6 µg DNA were added and the solution mixed briefly. The DNA/FuGene/medium mixture was incubated for another 30 min at room temperature and then was added to culture dish with the cells that had been washed and refed with 5 ml medium; the next day the medium was replaced.

3.5.5.2. Transfection with TransPass D1 (New England Biolabs GmbH, Frankfurt am Main)

One day before the transfection $4.5x10^5$ MCF-7 cells were seeded in a 3.5 cm culture dish. TransPass D1 transfection mixture was prepared according to manufacturers instructions. Briefly, for a 3.5 cm cell culture dish, 12.5 µl TransPass D1 were added to 2.3 ml medium without serum containing 5 µg DNA, mixed and incubated for 30 minutes. Cells were washed with serum free medium and incubated with the transfection mixture for 3-4 hours at 37°C and 95% humidity in the presence of 5% CO_2. Afterwards the transfection medium was replaced with supplemented medium.

3.5.6. Clone production

Transfected or transduced MCF-7 cells growing as isolated clones after selection could be picked from the culture dish. Therefore, the medium was removed, the cells washed with PBS and a thin film of PBS left on the cells. Single cell clones were dislocated mechanically with a pipette tip and introduced individually into wells of a 24 multiwell plate containing between 150-200 µl 0.05 % trypsin solution (Gibco, Invitrogen GmbH, Kalrsruhe). Cells were trypsinized for 5 min at room temperature and the reaction was stopped with 600 µl medium containing serum. Clones were grown at 37°C and 95% humidity in the presence of 5% CO_2.

3.5.7. Cell proliferation assay

Cells were plated onto 24 multiwell cell culture dishes in growth medium and allowed to attach for 24 hours. The cell proliferation kit II (XTT) assay (Roche, Mannheim) was used following the manufacturer's instructions. Briefly, for a 24 multiwell culture plate 3.4 ml of the XTT labeling/electron coupling mixture were added to 6.9 ml cell culture medium. Cells were rinsed twice with PBS and then incubated with 450 µl of the XTT/medium solution for 4-16 hours at 37°C and 95% humidity in the presence of 5% CO_2. After this incubation, the absorbance at 492 nm was measured using a microplate reader (Tecan, Crailsheim), the reference wavelength was 650 nm.

For cell proliferation or toxicity experiments $1x10^4$ and $1.8x10^4$ MCF-7 cells respectively, were used.

3.5.8. Immunofluorescence

Correct expression and subcellular localization of over-expressed, epitope tagged proteins were analyzed by immunofluorescence. Cells were grown to 70-80% confluence on gelatin coated glass coverslips, which were prepared by covering them with 0.1% gelatin in water (w/v) for 2-4 min directly before seeding the cells. Following fixation with 4% paraformaldehyde in 1xPBS (w/v) for 20 min at room temperature, cells were rinsed three times with PBS and permeabilized with 0.5% Triton X-100 in PBS for 15 min. After three washes with PBS, unspecific antibody binding was blocked by incubation with 2% BSA in PBS (w/v) for 20 minutes at 37°C. Incubation with the first antibody was done for 1 hour at 37°C in a humidified chamber. The unbound antibody was removed by three washes with PBS. The secondary antibody, conjugated to a fluorescent dye, was then added and incubation was continued for one hour at room temperature in a dark humidified chamber. After washing as before, nuclear DNA was stained for 15 min with DAPI (10 µg/ml in PBS). Finally, samples were washed again and mounted in Mowiol, which hardens overnight forming a permanent preparation. The distribution of the antigen was visualized under a fluorescence microscope (Axioplan 2 imaging, Carl-Zeiss AG, Göttingen).

Antibody	Commercial supplier	Working dilution
Cy3 conjugated anti-mouse	Dianova, Hamburg	1:1000 in PBS
Anti V5	Invitrogen, Karlsruhe	1:500 in PBS

Mowiol mounting medium:

13.3	%	Mowiol 40-88 (w/v)
33.3	%	glycerol (w/v)
133	µM	Tris base pH 8.5

Stir at 37°C overnight or until completely dissolved

3.5.9. Growth in soft agar

500 cells were suspended in 2xconcentrated supplemented DMEM and mixed 1:1 to 1 ml end volume with 0.7% bacto agar in water (w/v, Becton Dickinson, Heidelberg) and seeded into 3.5 cm culture dishes containing a 2 ml layer of solidified 0.7% agar in supplemented DMEM. Every 2 days 100 µl of DMEM containing 10% FCS was added. After growth for 2-3 weeks, medium was removed and a thin film of 0.005% crystal violet was added to the

plates and incubated for at least 4 hours to stain the colonies; those colonies containing more than 100 cells were counted.

Crystal violet solution:

0.005	%	crystal violet in 2% methanol (w/v)

Supplemented DMEM:

2	x	DMEM with phenol red
2	mM	L-glutamine
1	mM	Sodium pyruvate
100	U/ml	Penicillin G
100	µg/ml	Streptomycin
15	%	FCS (v/v)

3.6. Working with retrovirus

For this project self-inactivating, replication deficient, VSV-G pseudotyped retroviruses were used, which had to be handled and stored under biosafety level 2 conditions to comply with German laws concerning the work with genetically modified organisms.

3.6.1. Gene transduction with retroviral vectors

3.6.1.1. Production of replication deficient, VSV-G pseudotyped retroviruses

Replication-deficient (Miller, Skotzko et al. 1992; Boris-Lawrie and Temin 1993), Moloney Murine Leukemia Virus (MMLV)-based retroviruses were produced by transient cotransfection of 293T cells with a vector giving rise to the RNA to be packaged into the retroviral particles and two helper plasmids expressing the gag-pol and env-proteins (Markowitz, Goff et al. 1988). Briefly, $4x10^6$ cells were seeded in a 9 cm cell culture plate and allowed to attach for 24 hours. The next day cells were transfected with a mixture of 3.3 µg of the retrovirus construct, 5 µg pVpack-GP and 1.2 µg pMD-G using FuGene 6 (Roche, Mannheim) as described. Two days after transfection, the medium containing viral particles was collected and filtered through a 0.45 µm polyethersulfone filter. A concentrated polybrene stock solution was added to a final concentration of 5 µg/ml and the virus

suspension was shock frozen in liquid nitrogen and stored at –80°C. If necessary, the medium of the transfected cells was replaced for the collection of a second batch of virus supernatant on the next day.

Polybrene stock solution:

5 mg/ml polybrene in water

Filter sterilize, store at 4°C

3.6.1.2. Concentration of VSV-G pseudotyped MMLV-retroviruses

MMLV retroviral particles containing the Vesicular Stomatitis Virus G (VSV-G)-protein as envelope protein can be concentrated by ultracentrifugation or ultrafiltration. For this work concentration with Centricon Plus 20 ultrafiltration units containing a Biomax-100 filtration membrane, was chosen. 19 ml of virus containing culture supernatant were filled into the filtration units and centrifuged for 30 min at 1124 g and room temperature in a floor centrifuge. The filtrate in the collection tube was discarded and the filtration unit inverted. The concentrated virus was recovered by centrifugation for 1 min at 44 g and room temperature in the same centrifuge. To concentrate larger volumes of virus supernatant, several batches were concentrated as described, pooled and stored as above.

3.6.1.3. Infection of human cell lines with VSV-G pseudotyped retroviruses

For the infection with VSV-G pseudotyped retroviruses cells were seeded on regular culture dishes and allowed to attach overnight. Then the cells were washed with PBS and incubated for 5 hours with a virus suspension containing 5 µg/ml polybrene. After that, a twofold excess of medium was added and incubation continued overnight. The virus suspension was removed, the infected cells washed twice with PBS and refed with regular growth medium. If required, antibiotic selections were started two days after infection.

3.6.2. Generation of the switch reporter cell line

The switch reporter cell line was generated by retroviral infection at a very low multiplicity of infection to achieve single copy integration of the complete switch reporter cassette (Baer, Schubeler et al. 2000). MCF-7 cells were seeded at a density of $2x10^5$ cells in 9 cm cell culture plates and allowed to attach for 24 hours. Then the cells were incubated with 3 ml of serial dilutions (1:1 to 1:10.000) of the CMVlxtkneolxdnFADD virus particles as described in

section 3.6.1.3. Selection for transfected cells was begun on the next day after infection with 1 mg/ml G418.

3.6.3. Generation of the U3Cre integration library

3.6.3.1. Gene trap titration

The HeLa cell clone 260 (provided by J. Altschmied and K. Sturm) which carries the stably integrated switch reporter cassette PGKlxpurolxβgeo, was used for the titration. 3.5 cm diameter cell culture plates were seeded with $3x10^4$ cells and on the next day infected with 1 ml of a dilution series (0, 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, $10^{-5,}$ 10^{-6}) of the U3Cre gene trap virus on the as described in section 3.6.1.3. The next day the cells were collected and reseeded in 6 cm cell culture plates. Selection for Cre expressing cells with 1 mg/ml G418 was started two days after transduction. After 15 days of culture clones became visible, and were stained with crystal violet. Briefly, medium was removed from the plates, cells fixed for 2 min in cold methanol and a thin film 0.005% crystal violet in 2% methanol (w/v) was added and incubated for 10-20 min at room temperature until colonies were stained.

The colony numbers on the plates were:

Virus dilution	Colony number
0	full
10^{-1}	full
10^{-2}	284
10^{-3}	44
10^{-4}	0
10^{-5}	0
10^{-6}	0

Virus titer was derived by the number of colonies generated after G418 selection of HeLa reporter cell line infected with 800 µl virus solution. This colony number, which is indicative of virus integration into active genes, is multiplied by the average of U3Cre integrations in non active genes 100-200. The U3Cre virus titer estimation was $6x10^6$ virus/ml.

3.6.3.2. Creation of the gene trap library and selection of cell clones with integrations in TNFα inducible genes

Based on the dnFADD expression level after switch recombination and its response to TNFα, the switch reporter clone 103 was chosen for the establishment of the U3Cre gene trap library. $1x10^7$ cells were seeded at a density of $2x10^6$ in five 15 cm cell culture plates and infected with the U3Cre gene trap virus particles at an MOI of 0.5 ($1x10^6$ virus particles per dish).

Selection for cells, in which the switch reporter was not activated, was begun on the next day with 1 mg/ml G418. After 11 days of selection the cells were trypsinised and $2x10^7$ cells reseeded at a density of $2x10^6$ cells per dish in 15 cm cell culture plates. One day later 25 ng/ml TNFα was added to the dishes. Cells were treated for 7 or 3 days with TNFα and resistant clones were allowed to grow for 4-5 weeks.

4. - Results

4.1. Identification of TNFα induced genes in MCF-7 cells

4.1.1. Design of a cell line reporting TNFα induced gene trap insertions

In designing a reporter cell line capable of signaling TNFα induced gene trap insertions, two issues needed to be address. First, TNFα induces apoptosis in MCF-7 cells precluding a direct selection of productive gene trap events. Second, gene trap insertions into constitutively expressed "housekeeping" genes obscure the identification of regulated genes and require counter-selection. To address these issues the molecular switch vector shown in figure 7 (top) was cloned. It contains a puromycin resistance gene fused to an upstream phosphoglycerate-kinase-(PGK) promoter and a downstream transcriptional termination (polyA) site. In this vector, the puromycin-polyA cassette fulfils the role of a "STOP" cassette as it prevents the expression of a second gene inserted downstream encoding a N-terminally truncated FADD protein (dnFADD). dnFADD has been shown to exert a dominant negative effect on receptor mediated apoptosis including that induced by TNFα (Chinnaiyan, O'Rourke et al. 1995; Hsu, Shu et al. 1996). Since the puromycin-polyA cassette is flanked by *loxP* sites in direct orientation, Cre mediated recombination will excise it and thus position dnFADD immediately downstream of the PGK promoter. This means that in a cellular context the expression of puromycin resistance will be replaced by dnFADD which blocks TNFα induced apoptosis.

Figure 7 shows how reporter cells expressing the PGKpurodnFADD switch vector can be used in combination with a Cre recombinase transducing gene trap vector (i.e. U3Cre) to recover genes induced by TNFα. In a first step the reporter cells are infected with the U3Cre gene trap virus to produce an integration library with proviral insertions in a large collection of random chromosomal sites which can be either transcriptionally active (A) or inactive (B, C). Transcriptionally active sites induce Cre expression which recombines the switch. Thus, positive selection in puromycin eliminates the U3Cre insertions in constitutively expressed genes (A), like for example housekeeping genes. The surviving cells are now enriched for insertions into “silent” genes and can be used to identify genes that are activated by TNFα (B) by simply recovering the cells that became resistant to apoptosis induced by the cytokine.

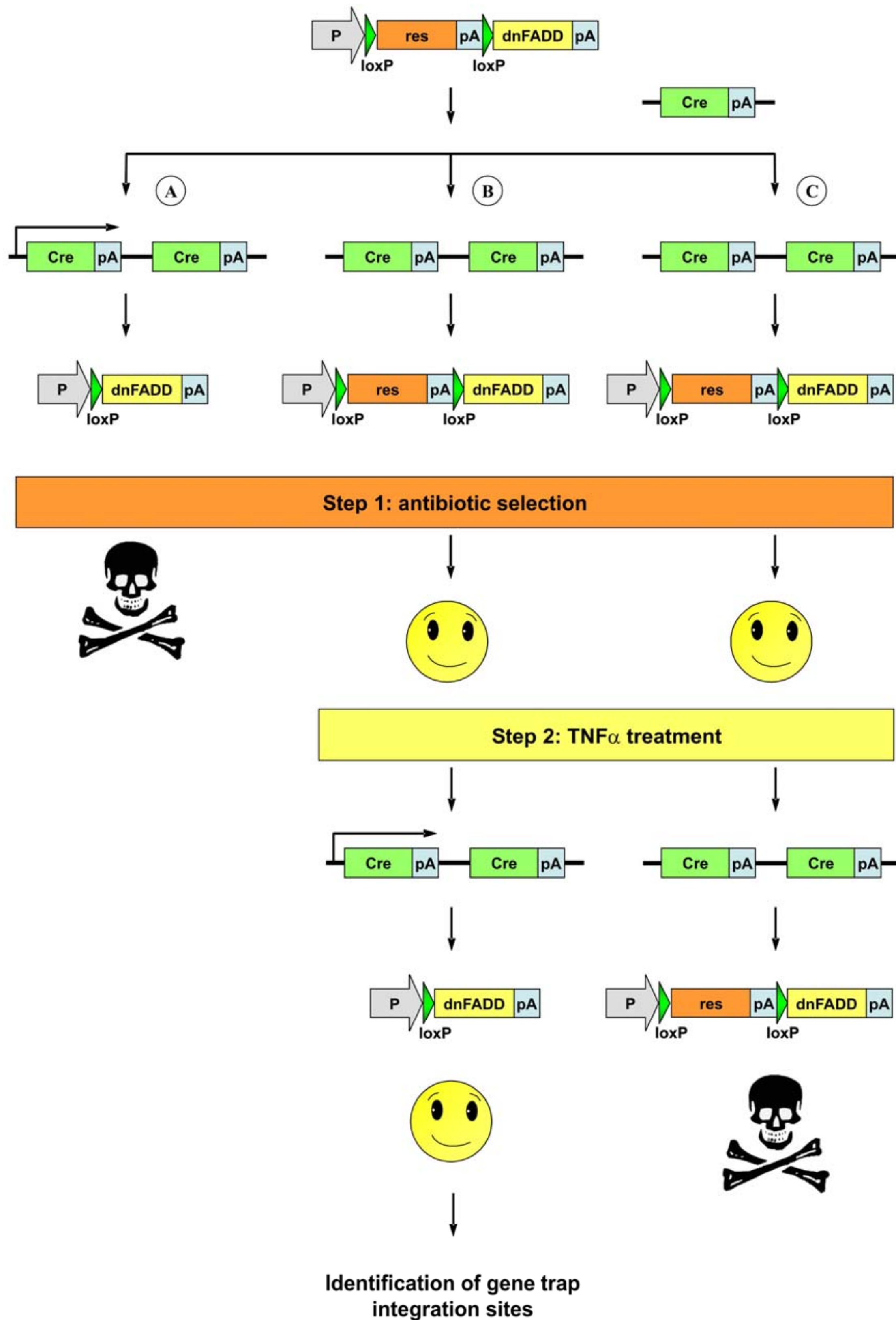


Figure 7. Selection scheme for the identification of TNFα induced genes. See text for explanations. dnFADD, dominant negative FADD; P, promoter; pA, polyA region; res, antibiotic selection cassette.

4.1.2. Stable over-expression of a dominant-negative FADD protects MCF-7 cells from TNFα induced cell death

As TNFα induced apoptosis was shown to be cell type specific (Fiers, Beyaert et al. 1999), it was essential to determine (i) the sensitivity of the MCF-7 cells against TNFα induced apoptosis and (ii) the efficiency with which *dnFADD* would block apoptosis in these cells. For this purpose pre-tested TNFα sensitive MCF-7 cells transfected with the plasmid pneoPGKlxpurolxdnFADD or its recombined derivative pneoPGKlxdnFADD (figure 8) were selected in G418 and exposed to 200 ng/ml TNFα for 6 days. Simple visual inspection quickly indicated that only some MCF-7 clones expressing the recombined plasmid survived in TNFα (data not shown), suggesting that dnFADD is expressed and functional in terms of apoptosis suppression.

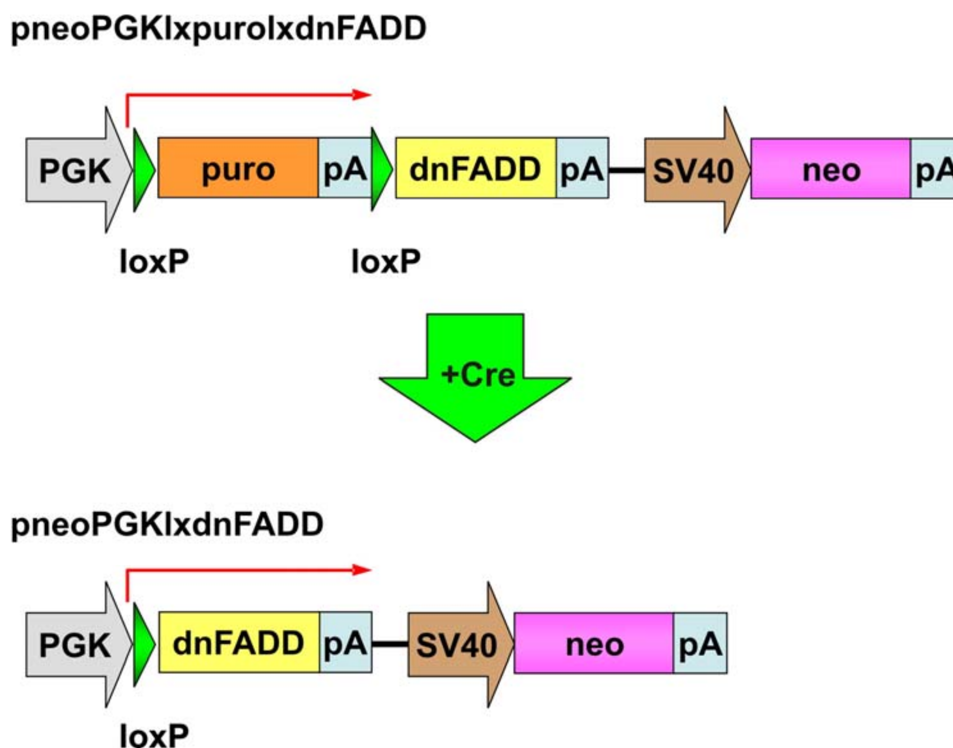
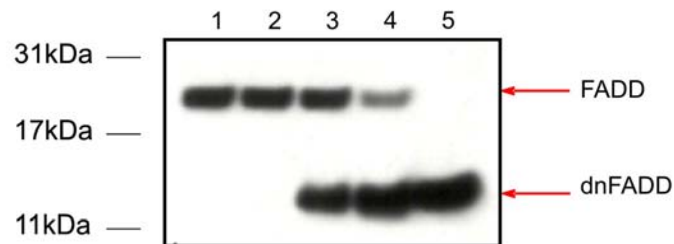


Figure 8. PGKpurodnFADD switch vector. In addition to a switch system with the mouse *phosphoglycerate-kinase* gene promoter (PGK) and the *puromycin-acetyl-transferase* gene (puro) which functions as a STOP cassette for the dnFADD cassette, the vector (PGKpurodnFADD) contains an independent selection cassette with the *neomycin-phosphotransferase* gene (neo) under the control of the simian virus 40 early promoter (SV40). The polyA (pA) regions are from the SV40 early region (for puro and neo) and the bovine growth hormone gene (for dnFADD). Cre induces recombination between the *loxP* sites and excises the puro-pA cassette to yield PGK-dnFADD.

A.-

1.- MCF-7
2.- PGK-puro-dnFADD
3.- PGK-dnFADD
4.- PGK-high dnFADD
5.- dnFADD



B.-

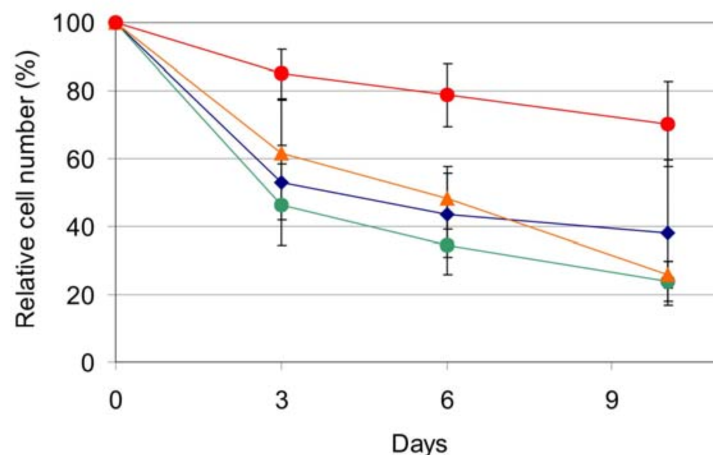


Figure 9. High level dnFADD expression confers TNFα resistance to MCF-7 cells. A.- Expression of dnFADD. MCF-7 cells were stably transfected with pneoPGKlxpurolxdnFADD (PGK-puro-dnFADD) or its derivative pneoPGKlxdnFADD (PGK-dnFADD). FADD levels in pools of transfectants were analyzed by Western blot using an antibody specific for human FADD recognizing the endogenous protein (FADD, 28 kDa) and the transgene encoded N-terminally deleted dominant negative form (dnFADD). PGK-high dnFADD denotes a PGK-dnFADD transfected cell population that had been pre-treated with 200 ng/ml TNFα for 6 days, dnFADD is a lysate from mouse cells carrying a transgene encoding human dnFADD (courtesy M.Zoernig, Georg-Speyer-Haus). 55µg of total cell lysates were loaded, with the exception of PGK-high dnFADD and dnFADD where 33µg and 3.6µg of cell lysate were used respectively. B.- High levels of dnFADD protect against TNFα effects. Cells (1.8×10^4 cells per well in a 24 well plate) were treated with 50 ng/ml TNFα for several days and the relative cell number was measured by assaying for metabolic activity with an XTT-assay. The curves show the mean and standard error from 3 independent experiments, represented is the relative cell number corrected for proliferation by comparing to untreated cells.

To investigate this further, levels of endogenous FADD and dnFADD were estimated in the transduced cells by Western blotting. Figure 9A shows that the ratio between dnFADD and endogenous FADD expression in PGKlxdnFADDSV40neo cells increased significantly after exposure to TNFα, suggesting that only cells with a vast excess of dnFADD over endogenous FADD are rescued from cell death (figure 9A). Moreover, cells expressing the non-

recombined plasmid failed to express dnFADD, indicating that its transcription is effectively blocked by the puromycin-polyA STOP cassette.

To further characterize the sensitivity to the cytotoxic action of TNFα, the stably transfected cell pools and MCF-7 wild type cells were treated with TNFα for 10 days and the cell viability was determined with an XTT assay (figure 9B). In contrast to wild type cells and the other pools, only cells expressing extremely high levels of the dnFADD protein were resistant to TNFα. These results provided evidence for a protective function of the dnFADD protein in MCF-7 cells suggesting that this cell line is amenable to the described gene trapping strategy.

4.1.3. Derivation of a MCF-7 reporter cell line suitable for trapping TNFα inducible genes

The previous experiments suggested that a high cellular dnFADD/FADD ratio is required for protecting MCF-7 cells from TNFα induced apoptosis. This is consistent with a dominant negative effect in which only a vast excess of the mutant dnFADD can completely block recruitment of caspases to the FADD-containing signaling complex. Consequently, to increase the conditional dnFADD expression from the switch vector, the PGK-promoter was replaced by a CMV-(cytomegalovirus immediate early region) promoter which has been shown to be highly active in mammalian cells (Foecking and Hofstetter 1986). In addition, the *puromycin-acetyl-transferase* gene was replaced with a HSV-*thymidine-kinase*/*neomycin-phosphotransferase* fusion gene (*tkneo*) to enable positive/negative selection (Schwartz, Maeda et al. 1991; Russ, Friedel et al. 1996) (figure 10). The previous experiments also indicated that the MCF-7 subline employed was suboptimal in its apoptotic response to TNFα. As it had been reported previously, MCF-7 sublines greatly vary in their apoptotic response to TNFα (Burow, Weldon et al. 1998). Therefore several other sublines were screened of which the MCF-7F cell line showed the fastest response to TNFα. Unlike the cells used for the initial experiments, 97% of MCF-7F cells were killed by TNFα within 3 days (figure 11B).

CMVtkneodnFADD switch vector was transduced into the genome of MCF-7F cells using a Moloney murine leukemia (MMLV) retroviral vector. Stable transformants were selected in G418 and expanded. In a first screen several clones were assayed individually for dnFADD

expression after Cre-induced recombination. Recombination was induced by transfecting a puromycin-selectable Cre expression vector into each single clone and by selecting in puromycin. Puromycin resistant clones were then analyzed for dnFADD expression by Western blotting. Figure 11A shows strong interclonal variation in dnFADD expression, most likely due to chromosomal position effects exerted on the switch vector integration site (Wilson, Bellen et al. 1990).

vCMVlxtkneolxdnFADD

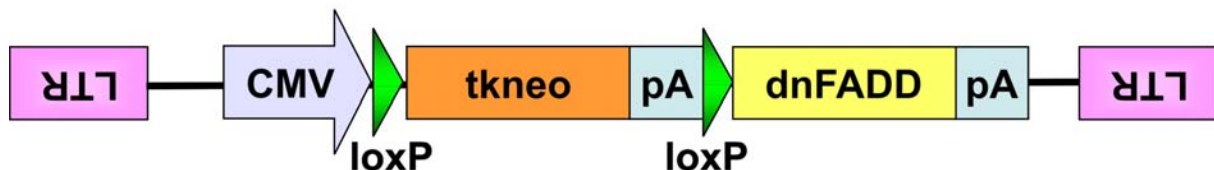


Figure 10. CMVtkneodnFADD switch vector. This vector contains the cytomegalovirus immediate early promoter (CMV), a floxed *tkneo* fusion and dnFADD cDNA, both with a polyadenylation region from the *bovine growth hormone* gene. The switch system is embedded into a Moloney murine leukemia virus backbone (symbolized by the long terminal repeats [LTR] of the proviral form) on the non-coding strand of the virus.

Based on the finding that a high dnFADD/FADD ratio is essential for mediating resistance to TNFα induced apoptosis (figure 9B), two clones (103 and 107) exhibiting high but different dnFADD/FADD ratios following recombination (figure 11A) were selected and assayed for their response to TNFα. For this, both recombined and non-recombined cells from each clone were treated with TNFα for 3 days. As expected, both non-recombined clones died within the first 48 hours of TNFα treatment. In contrast, the recombined clone 103 was completely resistant to TNFα induced apoptosis (figure 11B left). However, clone 107 while exhibiting some initial resistance to TNFα was still sensitive after prolonged exposure which clearly reflects its lower dnFADD/FADD ratio (figure 11B right). Based on these results clone 103 was selected for all further experiments.

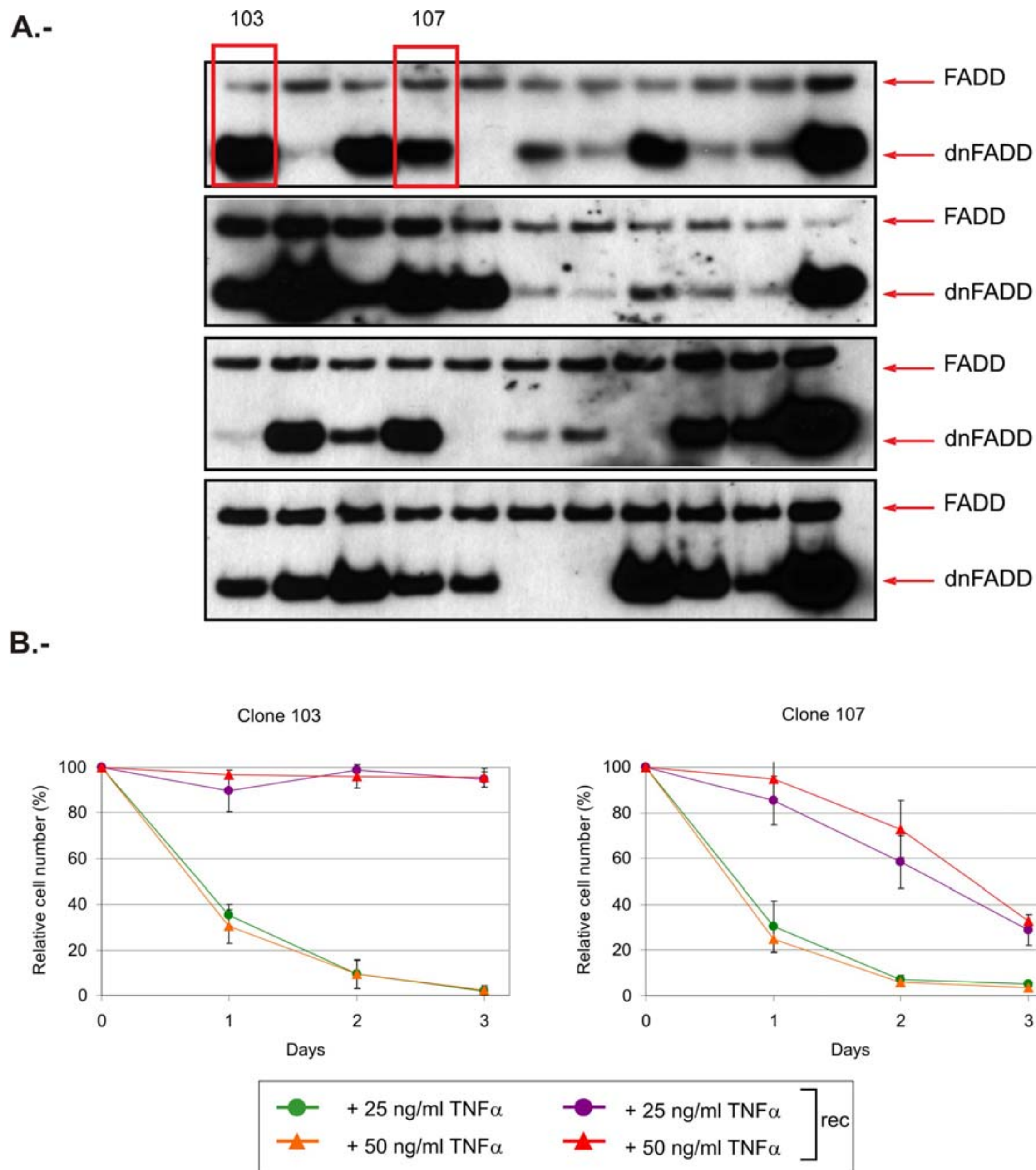


Figure 11. Characterization of MCF-7 switch reporter clones. A.- Expression of dnFADD after Cre-induced recombination. Cell clones stably transduced with the retroviral CMV-tkneo-dnFADD switch reporter were transfected with a puromycin-selectable Cre (pBabePGKcreSV40puro) expression vector and cultivated for 16 days in 1 µg/ml puromycin. Lysates (equivalent to 15 µg total cellular protein) were tested for dnFADD expression by Western blotting. Two clones with high dnFADD/FADD ratios are marked by red boxes. B.- TNFα sensitivity of selected switch reporter clones. Clones 103 and 107 ($1.8x10^4$ cells per well in a 24 well plate) were treated with different concentrations of TNFα for several days and the relative cell number was measured by assaying for metabolic activity with an XTT-assay. Rec denotes the cell pools, in which switch recombination had been forced by Cre expression. The curves show the mean and standard error from 3 independent experiments, represented is the relative cell number corrected for proliferation by comparing to untreated cells.

4.1.4. Isolation of cell clones with U3Cre integrations in TNFα inducible loci

After a reporter clone was established and characterized, a gene trap integration library for identification of TNFα inducible genes was generated. For a representative distribution of gene trap integration sites over the cellular genome, infection conditions were adjusted to obtain one gene trap insertion every 1000-5000 bp. Based on the size of a haploid human genome, $3x10^9$ bp, between $6x10^5$-$3x10^6$ independent integrations were necessary for saturating the genome with insertions at the desired density. This approximation is based on random retroviral integration, however, it has been reported that MMLV integrates preferentially into open chromatin regions, particularly into areas surrounding transcriptional start sites units (Wu, Li et al. 2003). Accordingly, complete gene coverage by retroviral gene trap vectors seems quite unlikely.

To obtain a library with the highest possible complexity and with mostly single copy integrations, 10^7 MCF-7 clone 103 cells were infected with U3Cre retrovirus at a multiplicity of infection (MOI) of 0.5. Under these conditions a theoretical infection efficiency of 100% would have generated an integration library with $2x10^6$ independent gene trap integrations. The infected cells were selected for 10 days in G418 (1 mg/ml), to eliminate integrations into constitutively expressed genes, as *Cre* expression should have lead to Cre-mediated excision of the tkneo cassette and thus loss of G418 resistance.

In a second selection step the G418 resistant population was treated with 25 ng/ml TNFα for 3 or 7 days. U3Cre integrations in TNFα responsive genes are expected to express Cre and induce recombination leading to the expression of dnFADD. This converts the cells to TNFα resistance and enables their selection. Parallel treatment of non-infected clone 103 cells served as a positive control for TNFα induced apoptosis. Most clone 103 cells were killed by TNFα within 3 days. In contrast, clones developed from the TNFα exposed gene trap integration library and 99 of them were used for further analysis.

PCR analysis of a representative fraction of these clones (28) could verify recombination in each case, indicating that TNFα resistance developed as a result of Cre-induced dnFADD expression (figure 12).

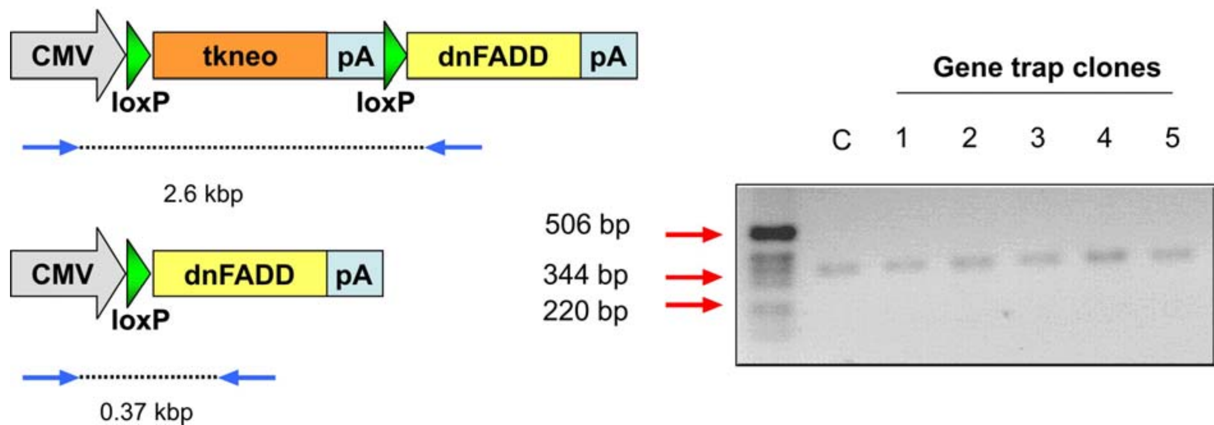


Figure 12. Gene trap clones show recombination in the switch system. The analysis is exemplified for 5 clones, genomic DNA from 5 cell clones obtained from the integration library after the two-step selection was isolated and analyzed by PCR using primers from the CMV promoter and the FADD coding region (CMVfor2/FADDrev4). An amplification product of 373 bp was indicative of Cre-mediated site specific recombination as the readout of gene trap activation. C is a derivative of the switch reporter cell clone 103, in which switch recombination had been induced by forced Cre expression.

4.1.5. Recovery of genomic gene trap sequence tags by inverse PCR

Genomic DNA sequences flanking the proviral integration sites (gene trap sequence tags, GTSTs) were retrieved by inverse PCR as shown in figure 13A. Amplification products obtained from 78 out of 99 clones were directly sequenced. Lack of amplification products from the remaining clones might have been the result of partial digestion of the genomic DNA and/or suboptimal circularization. Alternatively, high G-C content or extensive secondary structures are known to interfere with PCR amplification. Some clones generated more than one amplification product either due to oligoclonality or to partial cleavage of the genomic DNA. The latter was confirmed for several clones by the amplification of diagnostic intraviral fragments (figure 13).

Sixty nine out of the 78 recovered GTSTs were unique and exhibited typical cell-DNA-provirus junctions. As exemplified in figure 13B, GTSTs varied in size between 300-1500 bp.

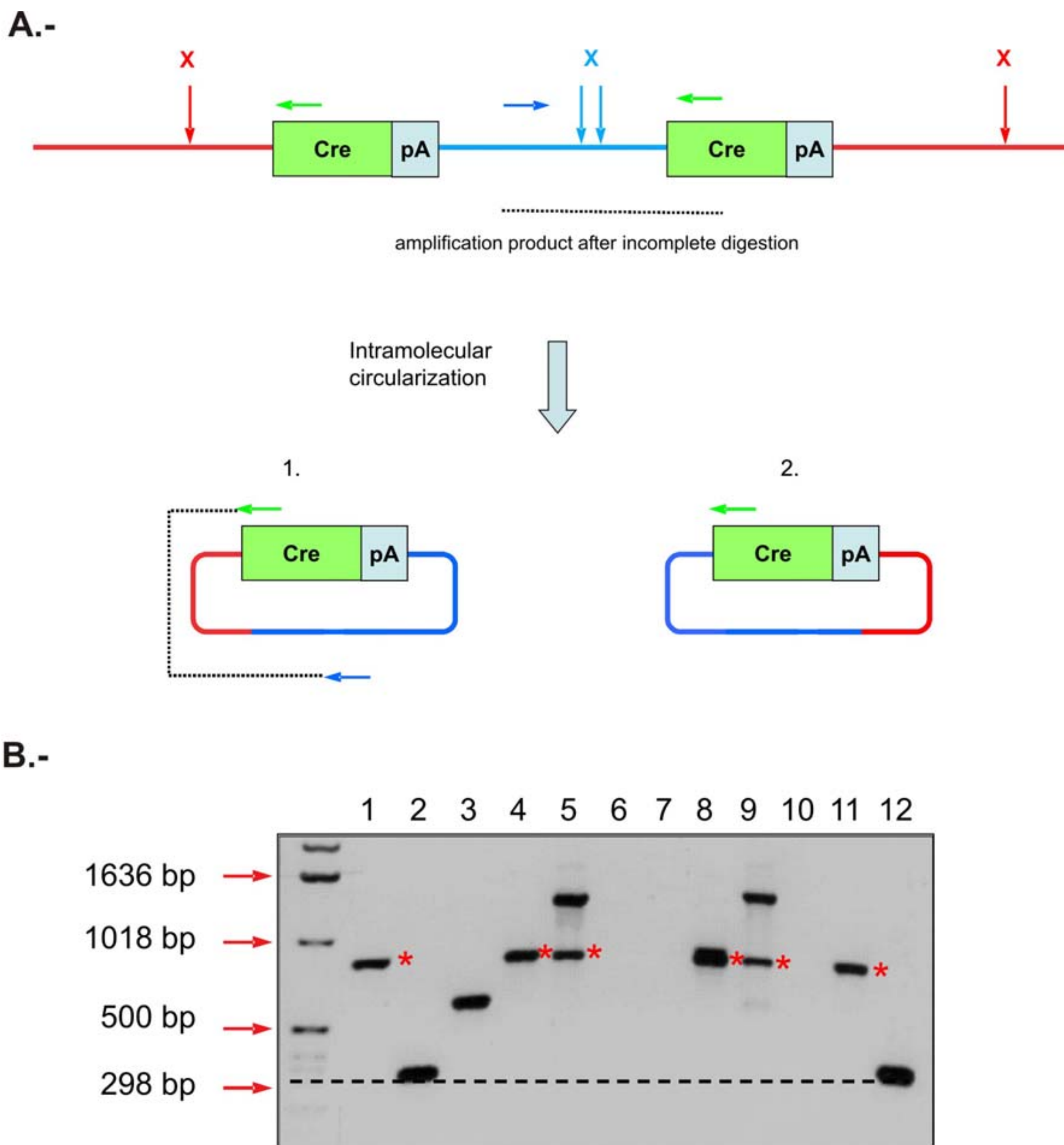


Figure 13. Inverse PCR isolates provirus-cellular junctions. A.- Reaction principle. Genomic DNA was digested with *Pst*I or *Nsp*I (shown as X), respectively and ligated under conditions favoring intramolecular circularization. The ligation products were amplified with the primer pairs Cre43/SY2 in the case of *Pst*I digestion and with CreiPCRrev2/iPCRU3 when *Nsp*I was used. Nested reamplifications were performed with Cre1P/SY1 (*Pst*I) and CreiPCRrev1/iPCRU4 (*Nsp*I). *Pst*I digestion and amplification with the indicated primers resulted in amplification of the junction between the 5'-LTR and the cellular DNA, whereas the combination of *Nsp*I with the corresponding primers yielded the 3'-junction. Primers are shown schematically as arrows. Fragments ≥310 bp (*Pst*I) and ≥362 bp (*Nsp*I) were indicative of successful amplification. Provirus-internal amplification products of 946 bp (*Pst*I) or 1522 bp (*Nsp*I) only occurred when restriction digests were not complete. B.- Analysis of iPCR products. Amplification products obtained from inverse PCRs of *Pst*I digested genomic DNA of 12 gene trap clones were separated by agarose gel electrophoresis. The dashed line indicates the minimum size expected and asterisks denote an undesired virus-internal amplification product.

4.2. *In silico* analysis of GTSTs

4.2.1. Identification of integration sites

After removal of their proviral segments, the 69 GTSTs were aligned to the human genome using the BLASTN algorithm and the NCBI (http://www.ncbi.nlm.nih.gov), ENSEMBL (http://www.ensembl.org) and Celera (http://www.celera.com) sequence databases.

GTSTs reported the presence of known transcripts, ESTs or genscans. Genscan is a probabilistic model of gene structure of human genomic sequences which incorporates descriptions of the transcriptional, translational and splicing signals as well as length distributions and compositional features of exons, introns and intergenic regions. The program can predict consistent sets of genes occurring on either one or both strands (Burge and Karlin 1997).

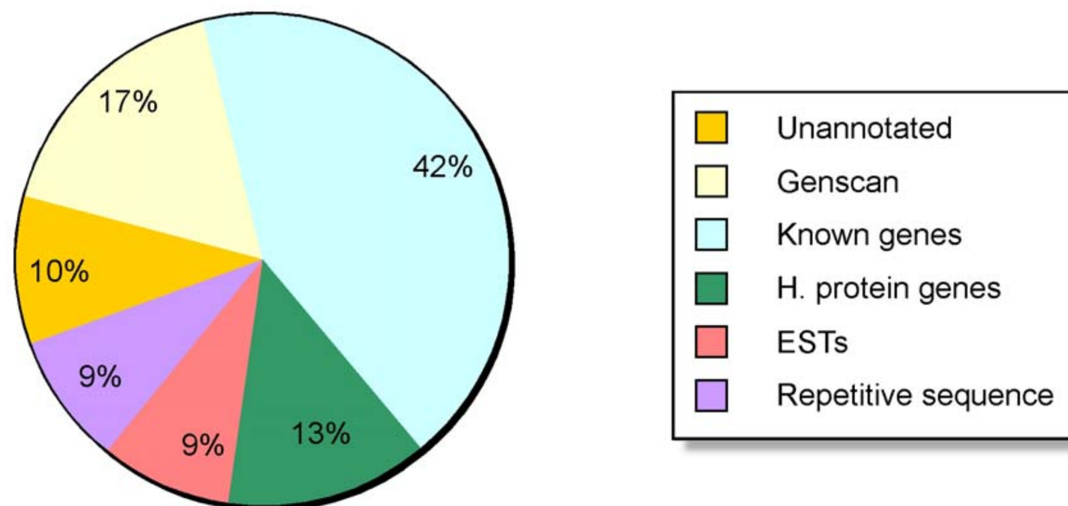


Figure 14. Bioinformatics analysis of gene trap integrations loci. BLAST searches with GTSTs from 78 gene trap clones retrieved integration into repetitive sequences, unannotated regions, genscans, known genes, hypothetical protein genes (known cDNAs with open reading frames encoding uncharacterized proteins) and ESTs.

Transcripts or genscans which were found within 2.5 kbp upstream or downstream of the gene trap integration were considered as potential TNFα regulated genes. Some integrations were between 2 genes which could share regulatory regions and thereby might be co-regulated by TNFα, therefore the number of recovered genes is higher than 69. As shown in figure 14, 35 (42%) GTSTs belonged to annotated genes, 14 (17%) to predicted (Genscan) genes, 11 (13%) to known cDNAs with open reading frames encoding uncharacterized

proteins (hypothetical protein genes), 8 (10%) to unannotated genomic sequence, 7 (9%) to ESTs and 7 (9%) to repetitive sequences. Integrations into unannotated regions were in some instances (3 out of 8 integrations) associated with transcription start site predictions by Eponine or First EF which could indicate the presence of a transcript.

Interestingly, 2 recovered genes showed more than one independent integration event (2 for *ESR*, 4 for *RACK7*), corroborating in this way Cre induction through TNFα in these genomic regions. These genes could represent hotspots for retroviral integration, but, due to the small sample size, this cannot be stated with certainty.

4.2.2. U3Cre insertions are mainly in 5'-introns

In accordance with the well known preference of retroviral integrations into the 5'-end of genes (Wu, Li et al. 2003; Bushman, Lewinski et al. 2005), most of the U3Cre insertions into annotated genes were near their 5'-end. An additional factor likely to affect U3Cre's preference for 5'-ends is a stop codon 18 bp upstream of the Cre AUG imposing a strong selection for integration events in which *Cre* provides the first AUG in the resulting fusion transcript (von Melchner, Reddy et al. 1990).

Although U3 type gene trap vectors were conceived as exon traps and therefore do not have a splice acceptor, the majority (22 out of 46) of U3Cre insertions were in introns. This is in line with similar results obtained in high throughput gene trap screens in which over 80% of the U3 gene trap insertions occurred in introns rather than in exons (IGTC database/ http://www.genetrap.org). As has been shown recently, U3 gene trap vectors frequently activate cryptic splice sites upstream of the insertions site resulting in fusion transcripts that contain intron sequences of variable size (Osipovich, White-Grindley et al. 2004). Surprisingly, 16 integrations were found upstream of the annotated genes, suggesting the presence of not yet annotated 5'-exons or genes.

Interestingly, insertions into annotated genes were equally distributed between the coding and non-coding strands suggesting that the latter may have trapped non-coding RNA (table 1).

Integration	Up-stream	Exon 1	Intron 1	Intron X	Last exon	Down-stream	Total
Sense	3	2	8	4	0	4	21
Antisense	13	1	6	4	1	0	25
Total	16	3	14	8	1	4	46

Table 1. Localization of gene trap integrations up to 2.5 kbp upstream/downstream and within known genes and hypothetical protein genes. X represents all introns except intron 1.

Integration \ Orientation	Sense	Antisense	Total
Known genes	15	20	35
H. protein genes	6	5	11
ESTs	6	1	7
Genscans	9	5	14
Total	36	31	67

Table 2. Orientations of gene trap integrations in recovered genes. Shown are the integrations in known genes, hypothetical protein genes (H. protein genes, known cDNAs with open reading frames encoding uncharacterized proteins), ESTs and Genscans. *ESR1-005* and *PRKCBP1*, both of them known genes, have 2 and 4 independent integrations respectively.

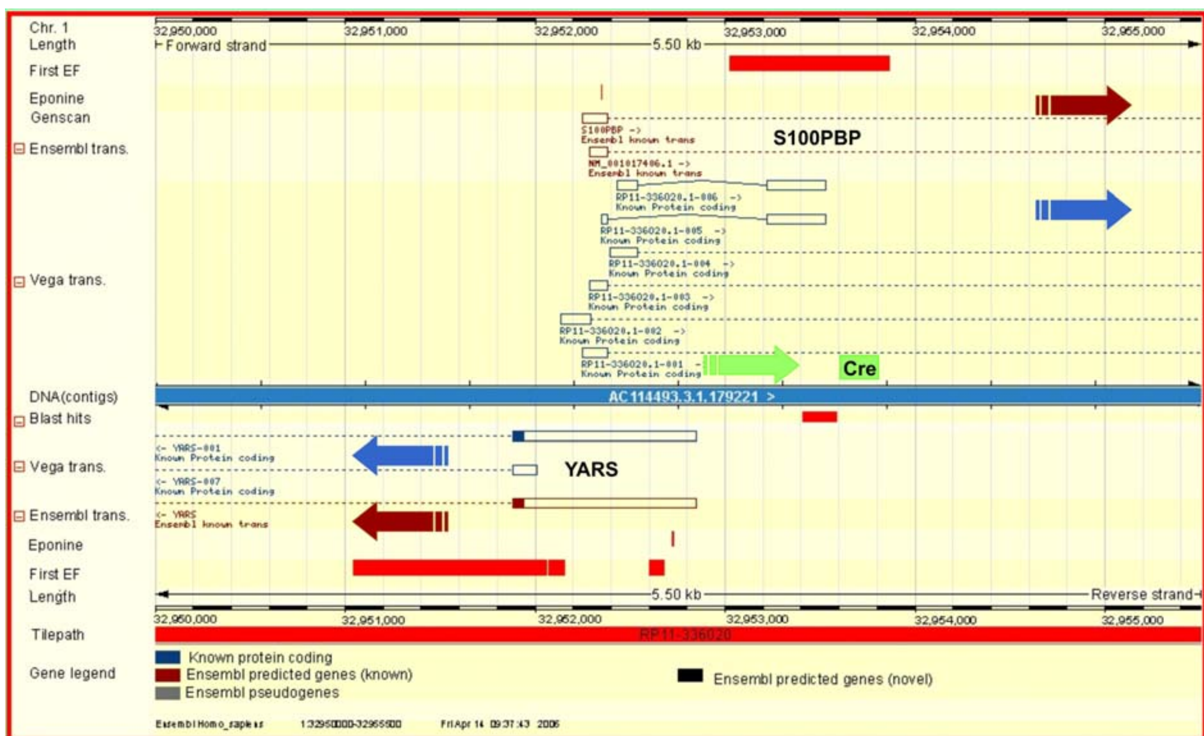
4.2.3. U3Cre traps putative antisense (non-coding?) transcripts

Data base analysis revealed that more than 40% of the U3Cre integrations were in antisense orientation relative to the transcriptional orientation of the trapped genes. This percentage was similar for known genes and hypothetical proteins (table 2). Since the gene trap strategy described here is very sensitive, as only a few Cre molecules per cell are required for recombination (Guo, Gopaul et al. 1997), it seems quite suitable for detecting weakly expressed genes. These include non-coding RNAs and natural antisense transcripts (NATs) which are missed by most functional genomics approaches because they are focused on

protein coding genes. Indeed, some of the antisense insertions into annotated genes corresponded to regions for which overlapping antisense ESTs or Genscans have been identified (figure 15A). In other instances, insertions occurred in regions, where sense/antisense gene pairs less than 1000 bp apart suggested the presence of bidirectional promoters (figure 15B).

Figure 15A.-

S100PBP and YARS



S100A10 and Genscan

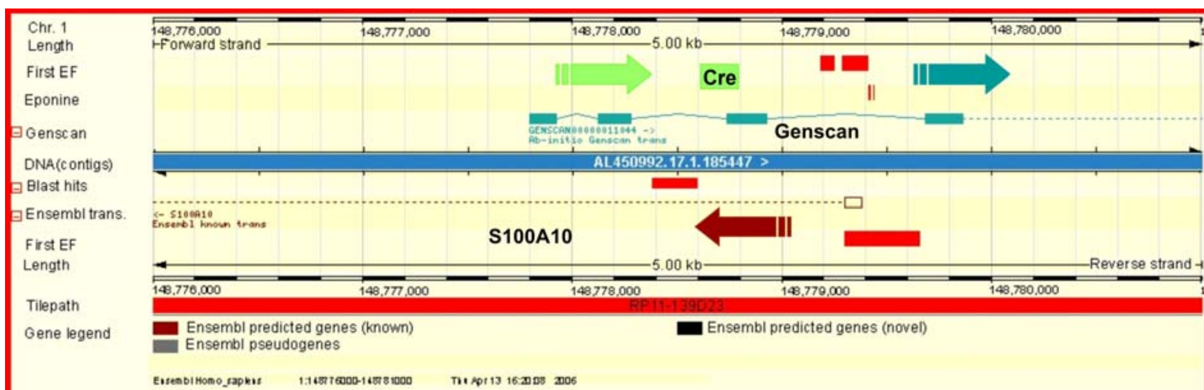
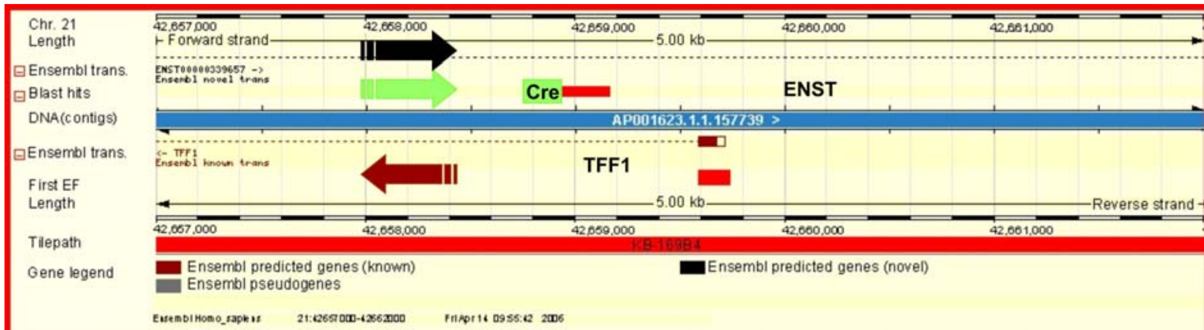
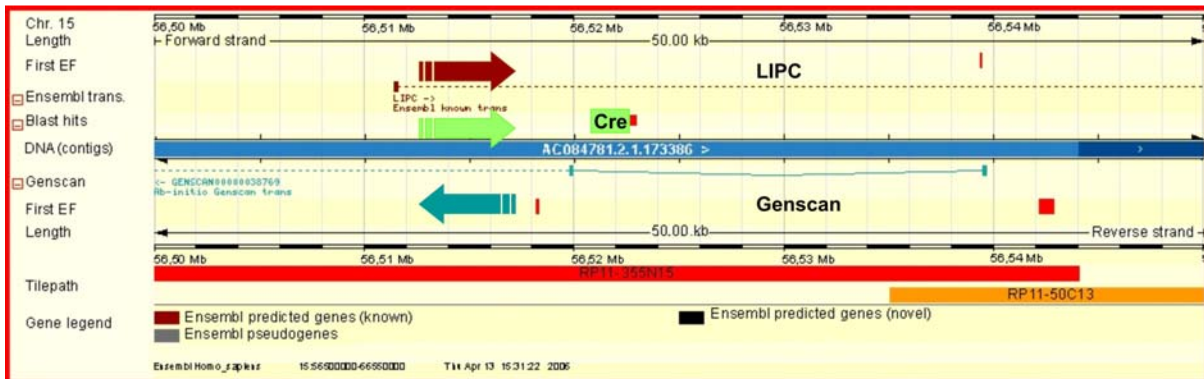


Figure 15A.-

TFF1 and EST



LIPC and Genscan



GRHL3/SOM/TFCP2L and Vega transcript

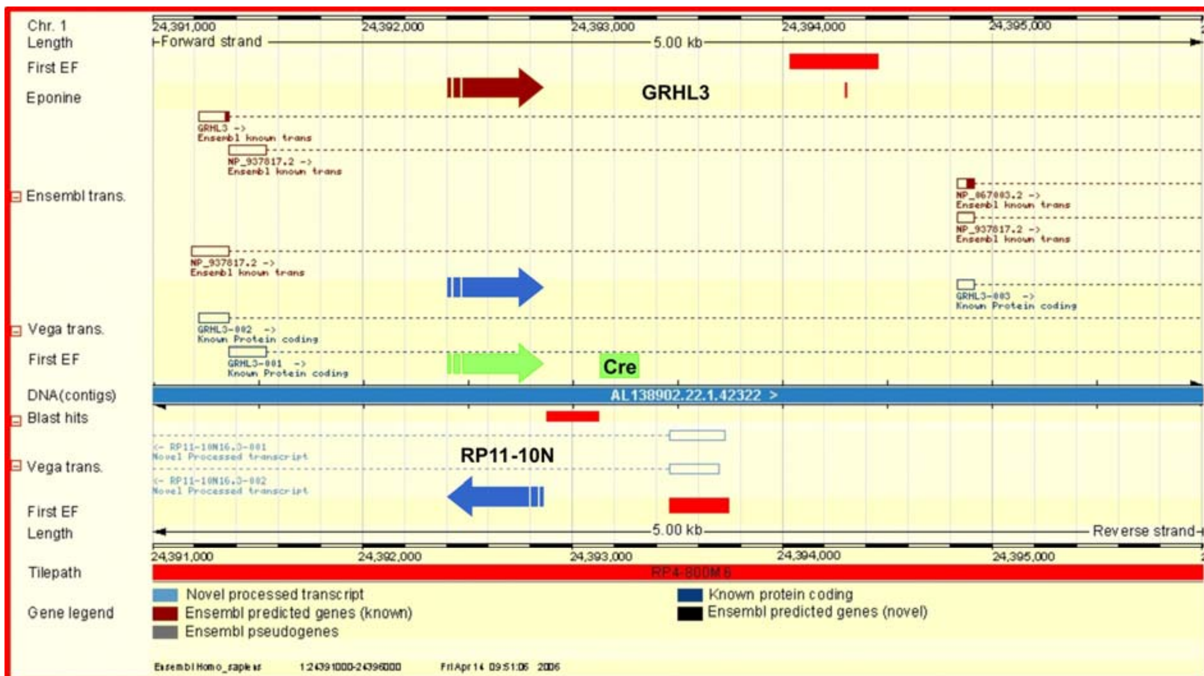
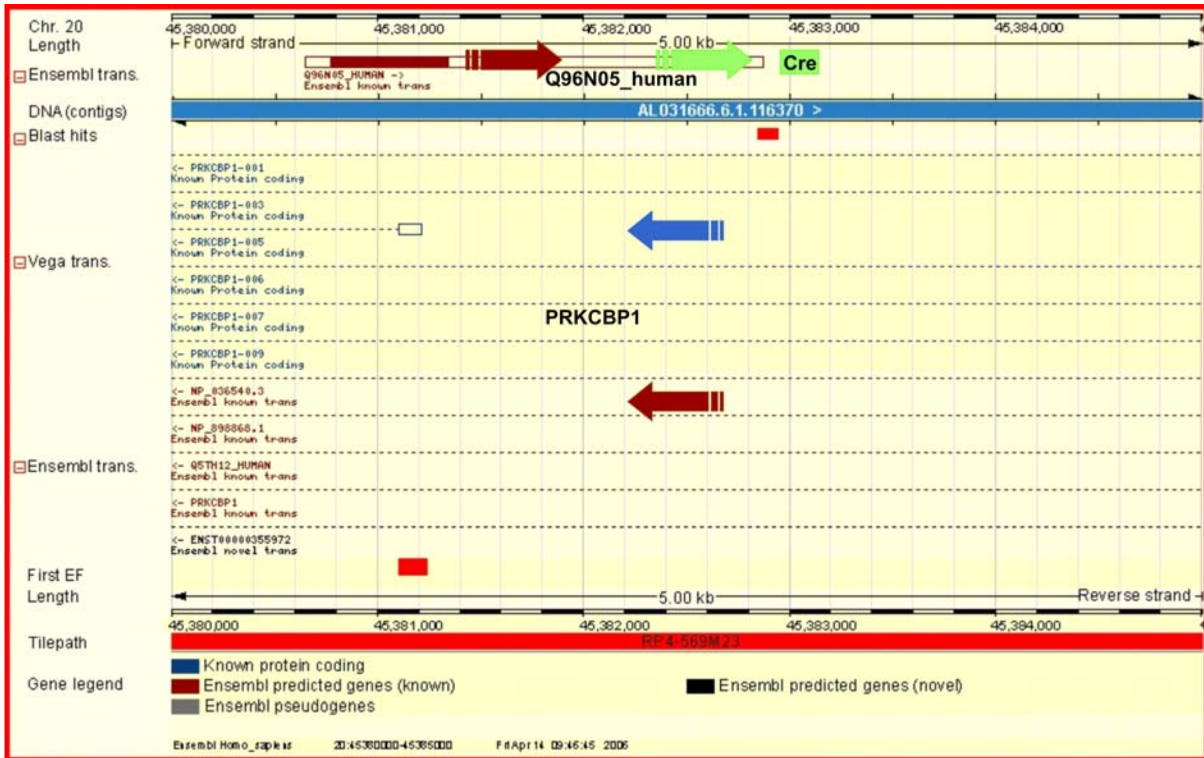


Figure 15A.-

PRKCBP1/Rack7 and Q96N05_human



CTNND2 and Genscan

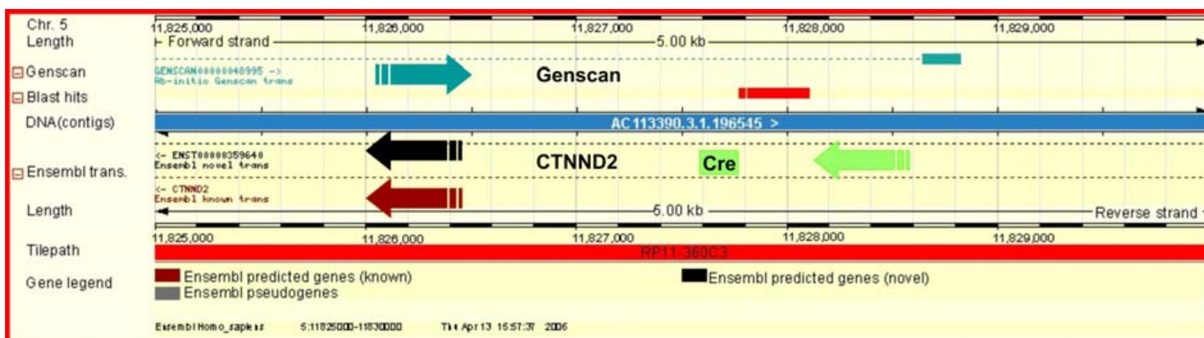
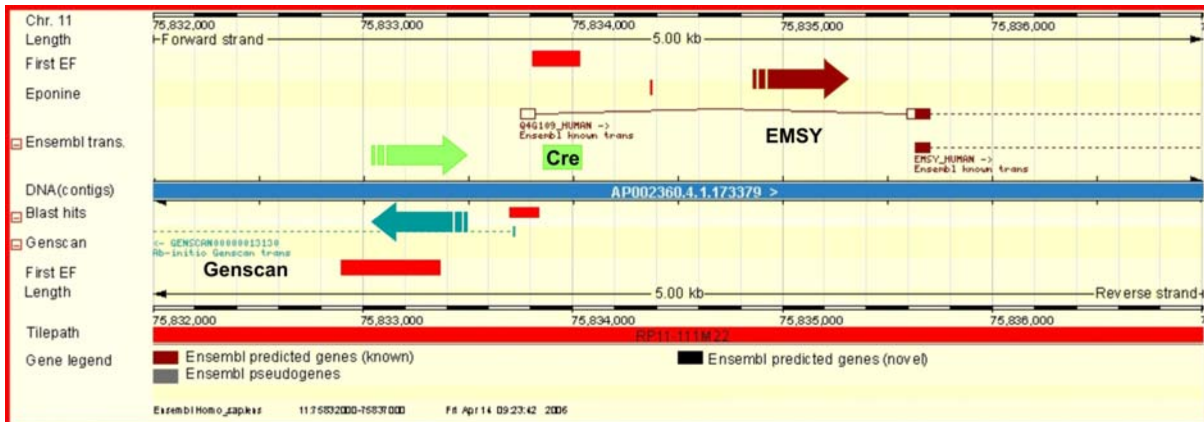
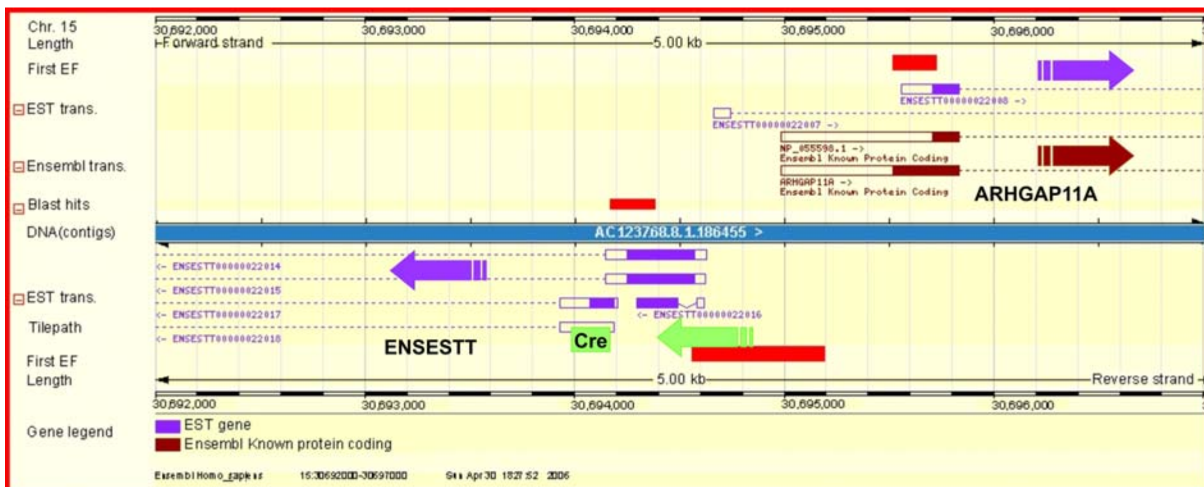


Figure 15B.-

EMSY and Genscan



ARPHGAP11A and ENSESTT



MBD4 and WDR10/IFT122

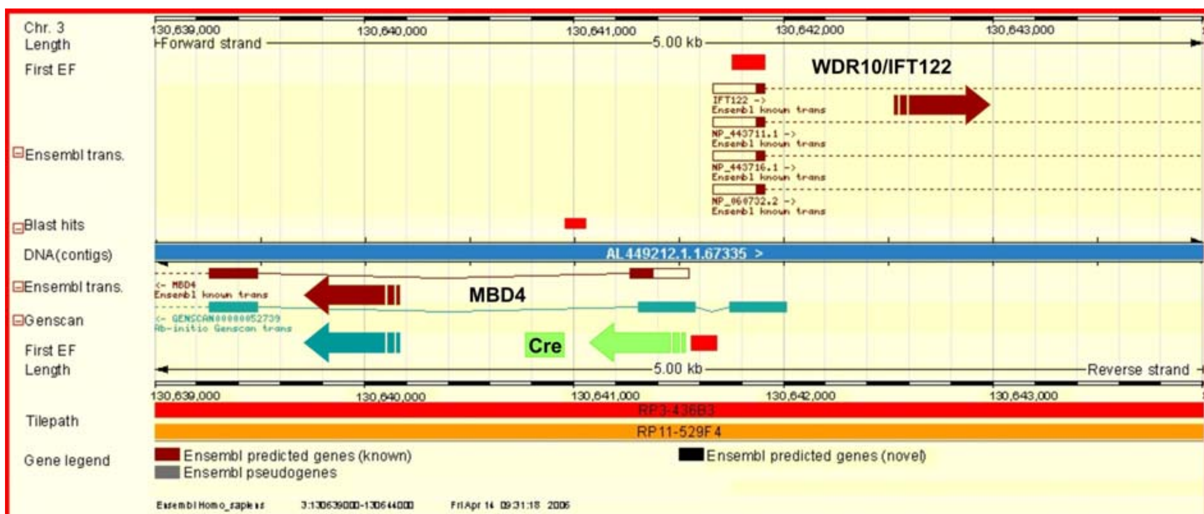


Figure 15B.- SLC12A2 and ENSTT

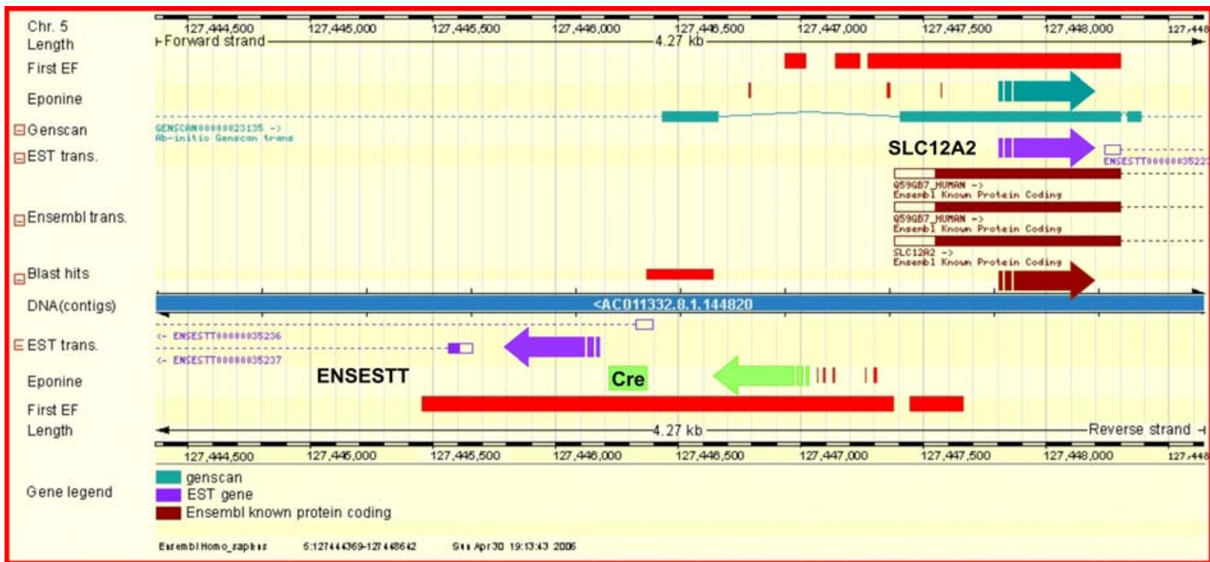


Figure 15. Gene trap integrations into regions with sense/antisense gene pairs. Depicted are the regions around the integration sites. Exons are shown as boxes, introns as dashed lines, solid arrows indicate the direction of transcription, with colors matching the corresponding transcripts. Blast hit denotes the identity between the sequence obtained in the inverse PCR and the genomic sequence. The position of the gene trap integration is indicated by a green box identified as Cre. First EF and Eponine are predicted first exons and transcription start sites, respectively. A.- Gene trap integrations which retrieved two overlapping transcripts oriented head to head. B.- Gene trap integrations which retrieve two transcripts oriented head to head with a maximum distance of 1kbp between their transcription start sites.

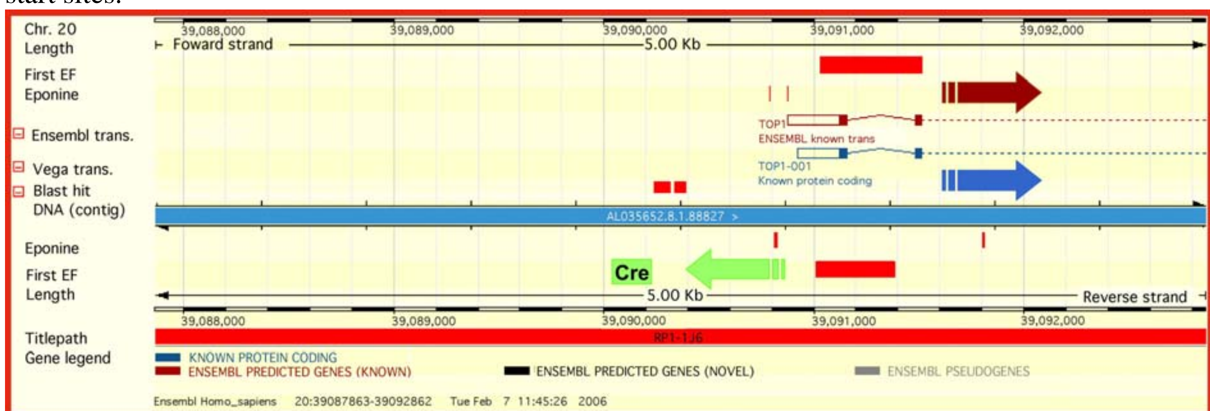


Figure 16. Transcription start site and exon prediction for an antisense gene trap insertion in the *topoisomerase 1* (*TOP1*) gene upstream region. Depicted are 5 kbp around the transcription start site of the *topoisomerase 1* gene. Exons are shown as boxes, introns as dashed lines, solid arrows indicate the direction of transcription, with colors matching the corresponding transcripts. Blast hit denotes the identity between the sequence obtained in the inverse PCR and the genomic sequence. The position of the gene trap integration is indicated by a green box identified as Cre. The *topoisomerase 1* gene is encoded on the upper DNA strand, whereas Cre is transcribed from the lower strand. FirstEF and Eponine are predicted first exons and transcription start sites, respectively. The later predict putative transcription start sites upstream of the gene trap, specifically in the second intron and in the promoter of the *topoisomerase 1* gene, moreover a first exon prediction is overlapping the *TOP1* exon 1.

Since U3Cre transcription from a non-coding strand of an annotated gene requires an RNA pol II promoter on this DNA strand, the regions upstream of the antisense insertions were screened for putative promoters using several *in silico* promoter prediction programs (Bajic et al., 2004; Rogic et al., 2002). The programs used were the dragon promoter finder version 1.5 (Bajic, Seah et al. 2002; Bajic and Seah 2003), the Neural Network Promoter Prediction program (NNPP, Reese 2001), First EF (first-exon and promoter prediction program for human DNA) (Davuluri, Grosse et al. 2001), McPromoter MM:II (The Markov Chain Promoter Prediction Server Massachusetts Institute of Technology Ohler, Harbeck et al. 1999), Promoter 2 (Knudsen 1999) and Promoter scan (Prestridge 1995).

Distance from U3Cre integration to predicted TSS (bp)					
NNPP	Promoter 2.0	Dragon Promoter finder	Mc Promoter MM:II	First EF	Promoter Scan
4810	-	-	-	-	-
4675	-	-	-	-	-
4140	-	-	-	-	-
3924	-	-	-	-	-
3332	-	-	-	-	-
3277	-	-	-	-	-
3110	2700	-	-	-	-
1998	-	-	-	-	-
1888	-	-	-	-	-
-	1600	-	-	-	-
-	-	1207	-	1151	-
986	-	**934**	-	**952**	-
		744			-
587	-	**581**	-	-	**798 to 548**
	500		-	-	-
442	-		-	-	-
288	-	-	-	-	-
280	-	280	-	-	-

Table 3. Prediction of transcription start sites (TSS) and promoter regions with different programs. The table shows predicted TSS and promoters in a region 5 kbp upstream of the gene trap integration in the *topoisomerase 1* (*TOP1*) gene, where the trap is 565 bp upstream of the *TOP1* transcription start on the non-coding strand, using different freely available prediction programs (for details see materials and methods). The numbers represent the distance in bp from the U3Cre gene trap to the TSS, in the columns are the predictions from different TSS prediction programs. Putative promoter regions and TSS based on the selection criteria explained in the text are shown in bold.

The results obtained by such analysis are shown in table 3 using the antisense insertion upstream of the *topoisomerase 1* (figure 16, *TOP1*) gene as an example. While the different programs identified a variety of putative transcriptional start sites upstream of the insertion, only 2 sites were predicted by three different programs suggesting that they are real.

Gene	Gene trap integration site relative to gene	Predicted distance between TSS and gene trap (bp)	Localization of predicted TSS relative to gene
ARHGAP11	upstream	510-518	*ARHGAP11*, upstream
BCL9L	upstream	1782-1709	*BCL9L*, exon 1
EED	exon 1	151-157	*EED*, exon 1
		165-246	*EED*, exon 1
KCTD5	upstream	906-984	*KCTD5*, intron 1
PITPNM2	last exon	3958-3966	*Q8TEM4*, exon 1
Rack7/PRKCBP1	upstream	1029-969	*Rack7*, upstream
Rack7/PRKCBP1	upstream	1019-954	*Rack7*, upstream
RNF184	upstream	1135-1221	*RNF184*, intron 1
SLC12A2	upstream	929-1021	*SLC12A2*, exon 1
TFF1	intron 1	2503-2535	*TFF1*, intron 2
TOP1	upstream	935-979	*TOP1,* intron 1
		581-587	*TOP1*, upstream
WDR10	upstream	931-953	*WDR10*, exon 1
YARS	upstream	597-675	*S100PBP*, intron 1
ZFP67	upstream	1971-1903	*ZFP67*, intron 2
ZNF143	upstream	958-999	*ZNF143*, intron 1
		784-838	*ZNF143*, intron 1

Table 4. Putative transcription start sites upstream of gene trap integrations on the non-coding strand of genes. The table summarizes the analyses exemplarily shown for one gene in table 3. The numbers represent the distance in bp from the U3Cre gene trap to the predicted TSS. Listed are the regions upstream of the gene trap antisense integration where at least 3 transcription start site prediction programs or two such routines and a promoter search algorithm predict a transcription start within a window of 95 bp. The position of the U3Cre integration site and the localization of the predicted TSS relative to the genes are also listed. *Rack7* is represented in the table by two independent integrations.

Accordingly, the analysis of 25 antisense insertions in known genes or hypothetical protein genes revealed transcription start sites (TSSs) for 15 integrations, for some of them even 2 TSS were predicted: 4 were predicted upstream of genes, 5 in the first exon, 7 in the first intron and 2 in the second intron of the genes harboring the gene trap integrations on their non-coding strands (table 4). The *EED*, *TOP1* and *ZNF143* genes were the only examples for which three programs predicted two different TSSs. Overall, more than 50% of the gene trap integrations in antisense orientation to known or hypothetical protein genes were associated with a TSS on the non-coding strand of the known transcript. Most of these TSSs are localized either in the first exon or first two introns, substantiating the predictions made above about putative antisense transcripts.

4.2.4. Recovered genes belong to different functional gene classes

TNFα activates transcription of a variety of genes and evokes a wide spectrum of cellular reactions. Therefore it was not surprising that the trapped genes belonged to 9 different functional gene classes according to their gene ontology (GO) classification (table 5).

Category	Cases
Nucleic acid binding/ regulation of transcription	11
Transport	6
Protein synthesis-degradation-modification	7
Kinase and phosphatase	4
Cytoskeleton and basement membrane	2
Calcium binding	2
Metabolism	2
Cell cycle	1
None of these categories	6

Table 5. Gene ontology classification of trapped known and hypothetical protein genes.

The largest group (11) represented nucleic acid binding proteins and/or transcriptional regulators. Two other groups with 7 members each represented transport proteins and molecules involved in protein synthesis, modification and degradation.

Interestingly, some trapped genes (*AMPO*, *BCL9L*, *JunB*, *KNSL1/KIF11*, *ZFP67*), although belonging to different gene ontology classes, seem to be involved in related cellular processes, e.g. cell cycle regulation or extracellular matrix reorganization.

4.2.5. Recovery of cancer-related genes

Most of the recovered genes have not been previously linked to TNFα signaling. A likely reason for this is the sensitivity of the gene trap screen, which seems to recover genes that are not easily recovered by other methods because they are either too weakly expressed or only transiently induced. Nevertheless, some of the genes have been directly or indirectly associated with apoptosis or cancer progression in other systems as explicated below.

a) *JunB* codes for a component of the activating protein-1 (AP-1) transcription factor complexes, which are important in the control of cell growth, differentiation and neoplastic transformation. Mice lacking JunB expression in the myeloid lineage developed a myeloproliferative disease eventually progressing to blast crisis that resembles human chronic myeloid leukemia (CML). The immature and hyperproliferative phenotype of JunB deficient myeloid cells is fully reverted by ectopic expression of *JunB*. These results identified *JunB* as a key transcriptional regulator of myelopoiesis and a potential tumor suppressor gene (Passegue, Jochum et al. 2001). Passegue and colleagues showed that JunB over-expression decreased the frequency of long term hematopoietic stem cells (HSCs), while JunB inactivation specifically expanded the numbers of long term HSCs and granulocyte/macrophage progenitors, resulting in chronic myeloproliferative disorder (Passegue, Wagner et al. 2004).

The expression levels of JunB are significantly impaired in human CML cases due to the inactivation of the JunB gene by methylation (Yang, Liu et al. 2003). JunB is also under-expresed in the majority (73.3%) of hepatocellular carcinomas (Chang, Yeh et al. 2005).

b) The glycolytic enzymes hexokinase I and II (encoded by *HKI* and *HKII*) bind to the outer mitochondrial membrane (OMM) with high affinity. This association is mediated at least in part, by specific interactions with the OMM voltage-dependent anion channel (VDAC). Furthermore, hexokinase has been implicated in the regulation of VDAC opening, as VDAC assumes an "open" conformation when associated with mitochondrial hexokinase that permits adenine nucleotide exchange. Upon disruption of this interaction, VDAC adopts a closed conformation that prevents further exchange such that intramitochondrial ADP becomes limiting for oxidative phosphorylation (Robey and Hay 2005). Over-expression of *HKI* (Gottlob, Majewski et al. 2001; Bryson, Coy et al. 2002) and of its functional variant - *HKII* - (Majewski, Nogueira et al. 2004) in Rat-1 cells protects from growth factor withdrawal plus U.V. induced apoptosis.

c) Estradiol (E_2) is the main mitogen for normal breast epithelial cells. Its action is mediated by two receptors, one of which (ESRα/ESR1) is over-expressed in more than half of all breast cancer cases (Ali and Coombes 2000). In addition, estrogen has been shown to protect cancer cells against TNFα induced cell death (Burow, Weldon et al. 2001). This protective effect is associated with a translocation of p53 to the cytoplasm preventing it from transactivating cell death genes. However, it has been shown recently in MCF-7 cells that only ESRα activation is protective against TNFα. ESRβ activation increases the sensitivity of MCF-7 cells towards TNFα presumably by failing to translocate p53 to the cytoplasm (Lewandowski, Thiery et al. 2005).

d) EMSY is a protein that colocalizes with BRCA2 to chromosomal sites of DNA repair and interacts with proteins involved in the chromatin modeling. EMSY binds to the region of the BRCA2 responsible for transcriptional activation, an excess of EMSY silences this function of BRCA2. The physiological role for the BRCA2 transactivation domain remains uncertain. However, it has been shown that BRCA2 can stimulate transcription of androgen receptor regulated genes in cooperation with histone acetyltransferases. Whether EMSY over-expression affects transcriptional regulation, DNA repair or chromatin remodeling will require further studies.

The *EMSY* gene is amplified in 13% of sporadic primary breast cancers and 17% of ovarian cancers (Hughes-Davies, Huntsman et al. 2003; Brown, Irving et al. 2006).

e) The Methyl-CpG binding domain protein 4 (MBD4) is a mismatch repair protein which excises thymidine from GT mismatches in methylated regions of the chromatin. MBD4 interacts with the mismatch repair/tumor suppressor protein mutL homolog 1 (MLH1) which protects cells from apoptosis induced by DNA damaging agents. Moreover, MBD4 interacts with the Fas-associated death domain (FADD) protein, which is crucial for TNFα induced apoptosis (see introduction section 1.4.1.). Interestingly, MBD4 promoted the apoptotic response to DNA damaging agents (Screaton, Kiessling et al. 2003). Moreover, MBD4 is frequently mutated in human colorectal carcinomas, exactly in the region required for its interaction with the MLH1 and FADD proteins (Bader, Walker et al. 2000).

f) S100A10 is found in most cells bound to its ligand annexin A2 in the heterotetrameric $(S100A10)_2$-(annexin $A2)_2$ complex AIIt. S100A10 has been shown to regulate plasma membrane ion channels as well as cytosolic phospholipase A2. The heterotetrameric form of S100A10 is on the extracellular surface of many cells, where the S100A10 subunit functions as plasminogen receptor. S100A10 is down-regulated in human esophageal squamous cell carcinoma (Ji, Zhao et al. 2004), over-expressed in gastric cancer (El-Rifai, Moskaluk et al. 2002), breast carcinoma (Carlsson et al., 2005) and renal cell carcinoma (Teratani, Watanabe et al. 2002).

g) *BCL9L/BCL9-2*. The switch between β-catenin's adhesive and transcriptional functions is modulated by phosphorylation of β-catenin which favors BCL9-2 binding and precludes binding with α-catenin. Over-expression of BCL9-2 induces epithelial mesenchymal transition (EMT) of non-transformed cells and increases β-catenin dependent transcription. A reversion of this effect was obtained by siRNA mediated-downregulation of BCL9-2 in the colon cancer cell line SW480, which expresses high levels of this protein and β-catenin. Besides the induction of an epithelial phenotype, a translocation of β-catenin from the nucleus to the cell membrane and a drastic reduction in cell migration were observed (Kramps, Peter et al. 2002; Brembeck, Schwarz-Romond et al. 2004; Harris and Peifer 2005).

h) CTNND2 (δ-catenin) is an adhesive junction associated protein that promotes cell scattering and is exclusively expressed in the nervous system. The region on chromosome 5p harboring CTNND2 is frequently amplified in cervical cancer (Huang et al., 2006). Delta catenin is significantly over-expressed in prostate cancer compared to benign prostate

hyperplasia and is considered as a potential diagnostic marker for prostate cancer (Burger et al., 2002; Lu et al., 1999).

i) TFF1 is a small cysteine rich secreted protein that is expressed at high levels in malignant breast epithelial cells where *TFF1* expression is regulated by estrogen. In addition, TFF1 stimulates the migration of human breast cancer cells (Prest, May et al. 2002).

j) SPATA 5 is a member of the AAA-protein family (ATPase associated with diverse activities) which is over-expressed at the malignant conversion stage of carcinogenesis (Liu et al., 2000).

k) S100PBR is an interaction partner of S100P. Both of them are over-expressed in pancreatic epithelial neoplasia and pancreatic ductal adenocarcinoma compared to normal pancreata (Dowen, Crnogorac-Jurcevic et al. 2005).

l) *FGD3* is a member of the *FGD1* family; *FGD1* encodes a guanine nucleotide factor that specifically activates the Rho GTPase Cdc42. Like *FGD1*, *FGD3* stimulates fibroblasts to form filopodia, actin microspikes formed by Cdc42 stimulation (Pasteris, Nagata et al. 2000). *FGD3* has been reported to be up-regulated in follicular thyroid tumors carrying the *PA8-PPARγ1* translocation (Lacroix, Lazar et al. 2005).

m) KNSL1/KIF11 (Koller, Propp et al. 2006) and TOP1 (Mialon, Sankinen et al. 2005; Adams, da Silva et al. 2006) are targets for chemotherapeutic treatment of cancer.

In addition to the above mentioned genes and their protein products, several other genes identified in the current screen (*ABCC3*, *EED*, *LIPC*, *SLC1A2*, *SLC12A2*) show altered expression levels in tumors detectable by microarray analysis (data available online at http://www.oncomine.org) Overall, about 44% of the known genes (59% when including the microarray data from oncomine) recovered in this screen were directly or indirectly related to cancer suggesting that the gene trap strategy employed here is suitable for the identification of cancer relevant genes.

4.3. Validation of the TNFα inducible genes

4.3.1. Regulation by TNFα

To investigate whether the genes recovered in this screen are *bona fide* TNFα regulated genes, 24 genes were selected for transcriptional analysis in the parental MCF-7 cells. Since the TNFα induction of these genes might be only transient (see introduction section 1.4.2), RNAs were extracted from MCF-7 cells at different time intervals of TNFα treatment and hybridized on Northern blots to gene specific cDNA probes. Blots were exposed to Phospho(r)imager screens and signal intensities were measured using Image Quant QT software. Transcript levels were normalized to actin transcripts as exemplified in figure 17.

Table 6 shows that of the 24 genes analyzed, 5 showed no regulation, and 19 were either up- or down-regulated by a factor of at least 1.45 at some time point following TNFα exposure. The regulated genes can grossly be classified into several groups:

a.- Genes with sense integrations between coding exons indicating entrapment of the mRNA; regulation of more than 2 fold: *C1QTNF6*, *SOM/TFCP2L4*.

b.- Genes with antisense integrations indicative of non-coding regulatory transcripts; regulation of more than 2 fold: *FGD3*, *S100A10*.

c.- Genes showing weak up- or down- regulation (1.45 to 2 fold) with sense or antisense gene trap integrations: *ABCC3*, *c20orf142*, *EED*, *ESR1, HKII*, *S100PBP*.

d.- Genes, in which a transcript has been trapped outside the annotated gene region, where the sense orientation of the trap indicates the presence of a nearby alternative promoter; regulation of more than 2 fold: *Jun B*.

e.- Genes, in which the trap is located outside of the gene in antisense orientation; regulation of more than 2 fold: *RNF184*, *SLC12A2*, *TOP1*, *WDR10*.

f.- Genes showing weak up- or down- regulation, in which a sense or antisense transcript has been trapped outside the gene: *BCL9L*, *FLJ14451*, *KCTD5*, *ZFP67*.

Interestingly, as seen in table 6 the analyzed genes showed different regulation patterns: 2 genes were permanently up-regulated (*BCL9L, JunB*), 1 permanently down-regulated (*FGD3*), 7 transiently up-regulated (*ABCC3*, *ESR1*, *HKII*, *KCTD5*, *RNF184*, *SOM/TFCP2L4*,

TOP1), 9 transiently down-regulated (*C1QTNF6*, *c20orf142/CT142*, *EED*, *FLJ14451*, *S100A10*, *S100PBP*, *SLC12A2*, *WDR10*, *ZFP67*) and 5 not regulated (*CTNND2*, *TFF1*, *ZNF143*, *ZNHIT2*, *ZNRF1*). Some of the genes classified as transiently down- or up-regulated showed both phenomena within the time window of the analysis, this is the case for *FLJ14451*, *HKII*, *RNF184*, *S100PBP*, *SOM/TFCP2L4* and *ZFP67* (figure 18).

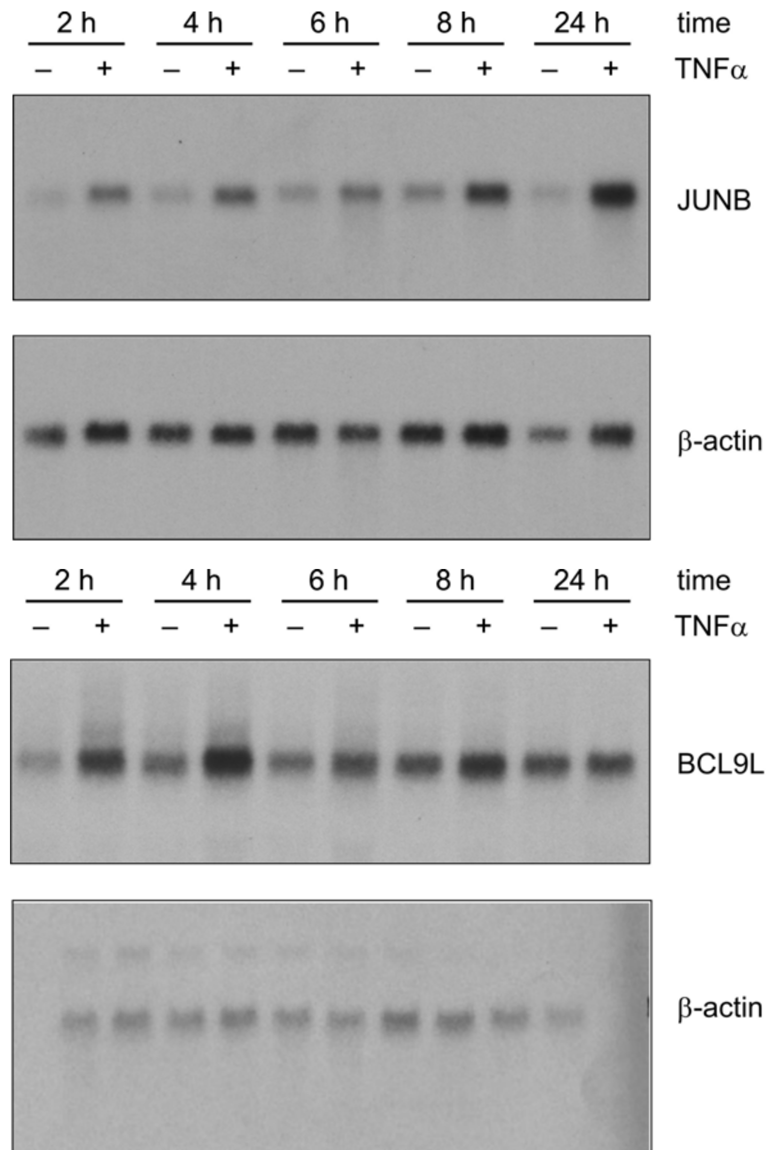


Figure 17. Selected examples of Northern blot analysis. MCF-7 cells were treated with 25 ng/ml TNF alpha (+) for the indicated times or left untreated (-) before RNA was extracted. 20 µg total RNA (blot hybridized with *BCL9L* probe) or polyA RNA extracted from 14 µg total RNA (blot hybridized with *JunB* probe) were separated on a formaldehyde-agarose gel and blotted onto a nitrocellulose membrane. The blots were hybridized with radioactive gene specific probes as indicated. Hybridization with the β-actin probe was performed for normalization.

Gene	Integration site	Orientation	Regulation factor at X hours after TNFα induction				
			2	4	6	8	24
C1QTNF6	intron 1	sense	1.16	1.03	1.61 ⇓	1.01	3.43 ⇓
SOM/TFCP2L4	intron 1	sense	1.93 ⇑	1.50 ⇑	1.64 ⇑	1.51 ⇑	5.63 ⇓
FGD3	intron 2	antisense	1.28	3.07 ⇓	4.85 ⇓	3.50 ⇓	3.34 ⇓
S100A10	intron 1	antisense	1.28	1.12	1.18	1.40	2.70 ⇓
ABCC3	intron 8	antisense	1.06	1.08	1.35	1.64 ⇑	1.02
c20orf142	intron 1	sense	1.25	1.02	1.16	1.17	1.47 ⇓
EED	exon 1	antisense	1.00	1.21	1.08	1.03	1.64 ⇓
ESR1	intron 1	antisense	1.38	1.83 ⇑	1.22	1.03	1.92 ⇓
HKII	intron 2	sense	1.14	1.31	1.63 ⇓	1.05	1.73 ⇑
S100PBP	intron 1	sense	1.58 ⇓	1.18	1.40	1.03	1.15
JunB	upstream	sense	1.67 ⇑	2.37 ⇑	1.90 ⇓	1.82 ⇑	3.33 ⇑
RNF184	upstream	antisense	1.57 ⇑	1.18	1.67 ⇓	1.12	3.40 ⇑
SLC12A2	upstream	antisense	1.08	1.12	2.36 ⇓	1.01	1.30
TOP1	upstream	antisense	1.01	1.98 ⇑	1.76 ⇑	1.45	2.11 ⇓
WDR 10	upstream	antisense	1.17	1.15	1.79 ⇓	1.01	2.61 ⇓
BCL9L	upstream	antisense	1.76 ⇑	1.80 ⇑	1.61 ⇑	1.48 ⇑	1.53 ⇑
FLJ14451	upstream	sense	1.95 ⇓	1.20	1.20	1.58 ⇓	1.03
KCTD5	upstream	antisense	1.56 ⇓	1.96 ⇑	1.93 ⇑	1.40	1.16
ZFP67	upstream	antisense	1.73 ⇓	1.15	1.08	1.08	1.51 ⇑
CTNND2	intron 1	sense	1.29	1.05	1.38	1.06	0.00
TFF1	intron 1	antisense	1.17	1.01	1.18	1.16	1.01
ZNF143	upstream	antisense	1.10	1.24	1.22	1.10	1.37
ZNHIT2	exon 1	sense	1.37	1.17	1.16	1.05	1.42
ZNRF1	intron 1	sense	1.14	1.07	1.28	1.38	1.08

Table 6. Regulation of 24 recovered genes. The table shows the U3Cre integration position, orientation and the regulation of mRNA levels by TNFα assessed by Northern blots and normalized for the amounts of β-actin RNA. Arrows indicate up ⇑ or down ⇓ regulation in comparison to untreated cells. Changes in RNA levels between 1.45 and 2 fold are shown in pink, changes of more than 2 fold in yellow.

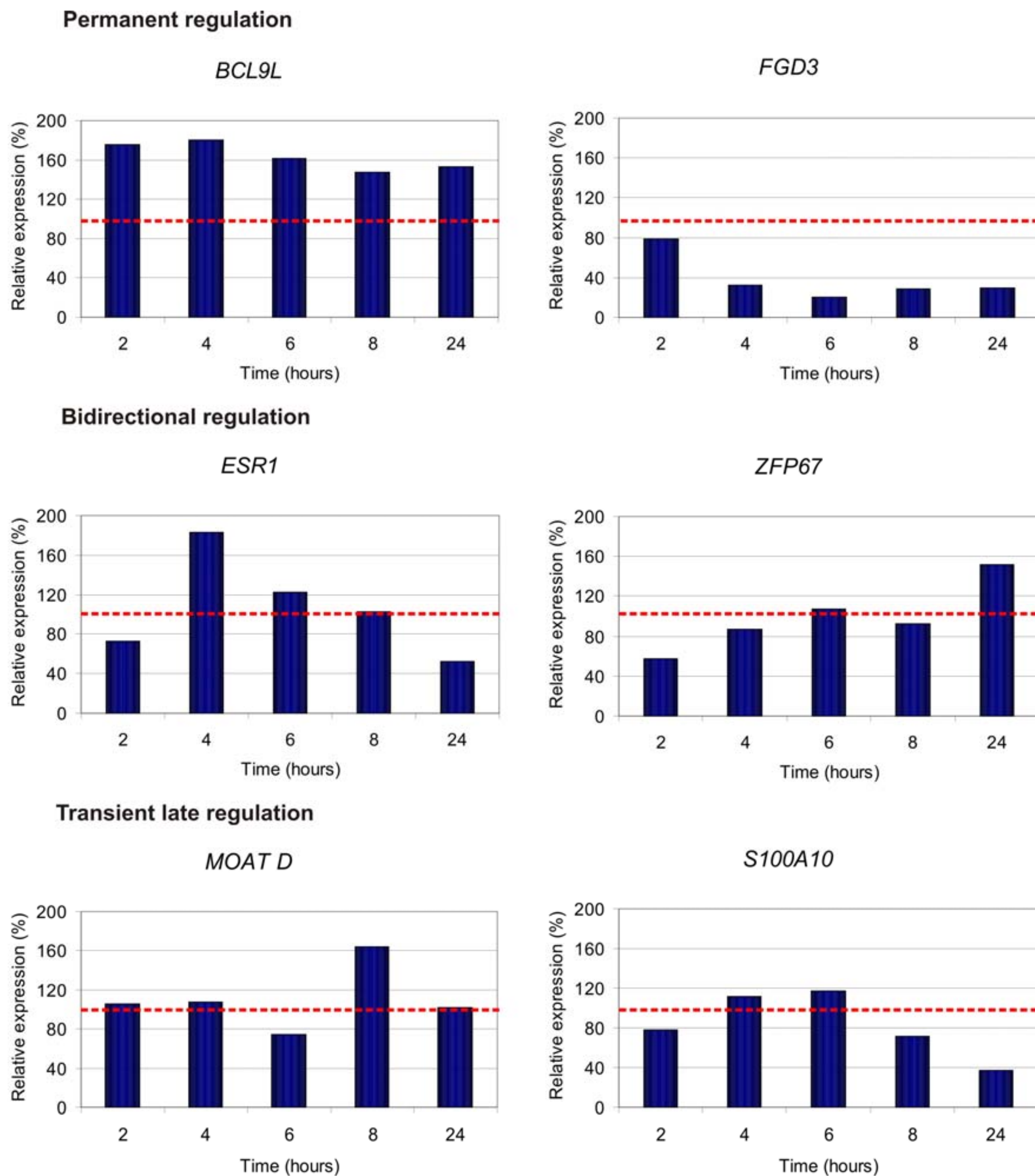


Figure 18. Genes exhibited different regulation by TNFα. Northern blots were quantified in a Phospho(r)imager and normalized for expression of the housekeeping gene *β-actin*. Bar graphs show the normalized RNA levels of TNFα treated cells. The expression level of untreated cells was always set to 100, which is indicated in the graphics by a dashed red line.

4.3.2. Selection of candidate genes for functional characterization

Candidate genes for functional analyses were selected based on their response to TNFα, the availability of full length cDNAs and a potential association with cancer. In accordance with

these criteria, two genes were selected; one encodes a known protein, whereas the other contains an open reading frame (ORF) for which no function had been assigned so far.

a) ZFP67/c-Krox belongs to a family of developmentally regulated genes. The first member of this family, *Krüppel*, was identified as a *Drosophila* segmentation gene (Schuh, Aicher et al. 1986). To date several homologues of *krüppel* have been isolated in vertebrates and are referred to as Krox genes. All these genes are early response genes which are transiently activated by serum or growth factors and some of them, such as Krox 20 and Krox 24 are expressed in a tissue specific manner (Lemaire, Revelant et al. 1988; Wilkinson, Bhatt et al. 1989). The protein encoded by *c-krox* has a POZ domain required for protein-protein interactions and a carboxy-terminal DNA binding domain consisting of krüppel-type zinc fingers. It acts as transcriptional repressors by recruiting histone deacetylases (HDACs) to chromatin.

ZFP67 is a key regulator of lineage commitment in immature T-cell precursors (Sun, Liu et al. 2005; Zamoyska 2005). More importantly *ZFP67* was shown to repress the expression of extracellular matrix proteins, particularly fibronectin (Widom, Culic et al. 1997). Inhibition of fibronectin in several *in vitro* transformation assays increased the rate of transformation, suggesting that the protein has a tumor suppressor function (Steel and Harris 1989).

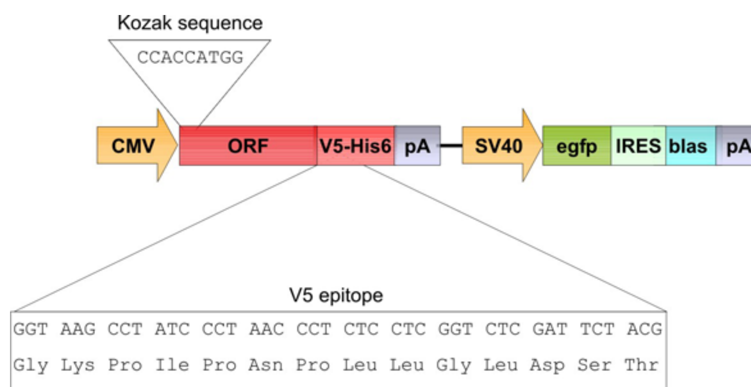
b) *FLJ14451* encodes for a hypothetical protein whose only obvious feature is the presence of 3 copies of a C-x8-C-x5-C-x3-H type zinc finger domain. Based on this, FLJ14451 might be related to a class of eukaryotic zinc finger proteins involved in cell cycle regulation. This class includes the human TIS11B (butyrate response factor 1 also known as EGF-response factor) protein, which is encoded by a member of the TIS11 family of early-response genes. TIS11B knockout mice die *in utero* as a result of abnormal placentation (Stumpo, Byrd et al. 2004). Another member of this family, the 35 kDa subunit of the human splicing factor U2AF was shown to play a critical role in mRNA by participating in 3' splice site selection.

4.3.3. Over-expression studies

To assess whether the selected candidate proteins are associated with the acquisition of oncogenic properties or other cellular alterations, they were over-expressed in MCF-7 cells. Towards this end, the ORF were amplified from the corresponding I.M.A.G.E. clones and

inserted into the expression vector pcDNA™6/V5-HisA downstream of a CMV promoter and in frame to a C-terminal V5/His_6 tag (figure 19A). To enable the identification of transfected cells, a second cassette expressing egfp and blasticidin-deaminase from an SV40 promoter was cloned downstream of the ORF expression cassette yielding the empty control vector pcDNA6egfpIRESblas and the two expression plasmids pCMVegfpIRESblasZFP67 and pCMVegfpIRESblasFLJ14451.

A.-



B.-

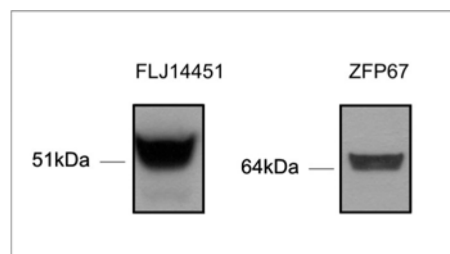


Figure 19. Over-expression of V5-tagged fusion proteins. A.- Expression constructs for candidate genes. Expression of the V5-tagged candidate proteins is driven by the CMV immediate early promoter (CMV). To create the fusion, the stop codon of the open reading frames (ORF) was removed. At the 5'-end a Kozak sequence is present for optimal translation. Outside the expression cassette the vectors contain the *enhanced-green-fluorescent-protein* gene (*egfp*) and the *blasticidin-deaminase* (*blas*) gene which provides blasticidin resistance under the control of the SV40 early promoter (SV40). The polyadenylation sequences for egfp and blas are derived from *bovine growth hormone* gene and SV40 early region, respectively. In the constructs for stable expression the egfpIRESblas cassette was replaced by neo. B.- Expression of fusion proteins. MCF-7 cells were transiently transfected with expression vectors for the candidate proteins, and total cell lysates (20 µg for ZFP67, 15 µg for FLJ14451) were analyzed in a Western blot using an anti-V5 antibody. The predicted molecular weights of the tagged proteins are 61.4 kDa (ZFP67) and 49.3 kDa (FLJ14451) respectively.

As blasticidin selection was not efficient for the selection of stably transfected MCF-7F cells, a second set of expression vectors was constructed, in which the egfpIRESblas cassette was

replaced by a *neomycin-phosphotransferase* gene driven by the SV40 early promoter (pcDNA6neo, pCMVFLJ14451neo, pCMVZFP67neo).

To verify the functionality of the expression plasmids, they were transiently transfected into MCF-7 cells and the respective proteins were visualized by Western blotting using an antibody against the V5 epitope tag. Figure 19B shows that in both cases the transfected MCF-7 cells expressed a protein of the predicted size, indicating that protein translation is intact.

4.3.4. Subcellular localization

As a first approximation to a function, the subcellular localization of the candidate proteins was assessed by immunofluorescence following transient transfection of MCF-7 cells with the construct pcDNA6egfpIRESblas and the expression plasmids pCMVegfpIRESblasZFP67 and pCMVegfpIRESblasFLJ14451 (figure 20).

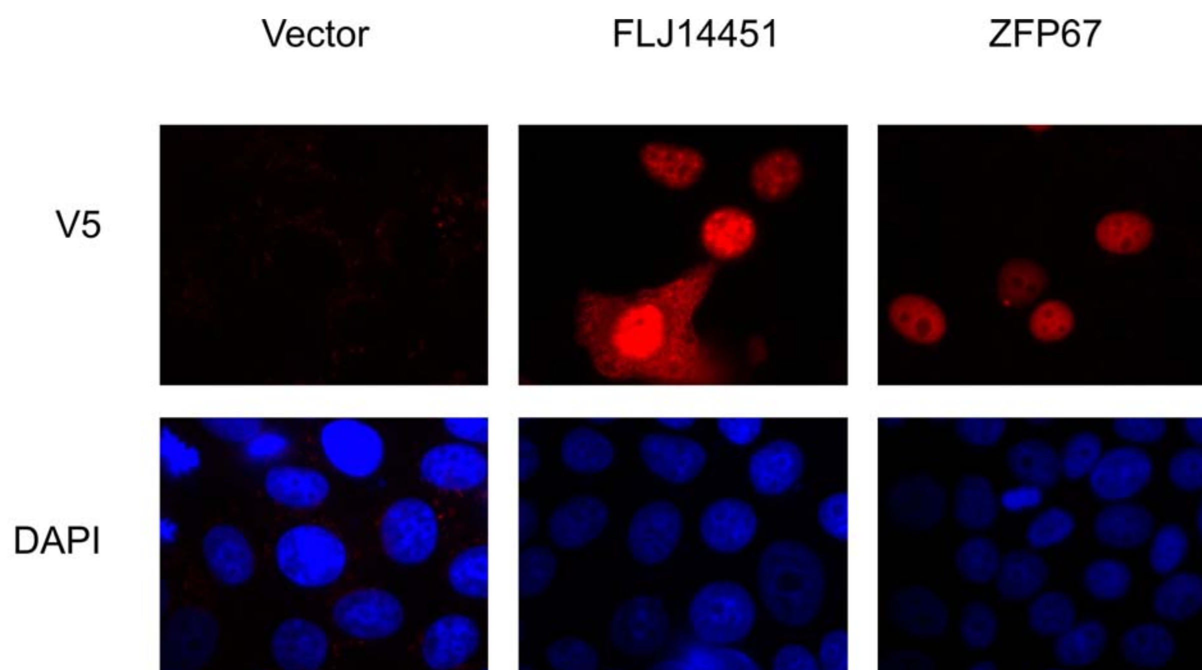


Figure 20. Candidate proteins are located in the nucleus. MCF-7 cells were transiently transfected with expression vectors for V5 tagged ZFP67 and FLJ14451. Proteins were detected with a mouse antibody against the V5 tag followed by a Cy3 coupled anti-mouse antibody (red); DAPI was used to counterstain nuclei (blue).

As predicted by its function, the transcription factor ZFP67 was found exclusively in the nucleus. Similarly, FLJ14451 was also mainly in the nucleus but unlike ZFP67 was concentrated in discrete speckles reminiscent of PML-nuclear bodies (PML-NBs; Maul, Negorev et al. 2000). An association of the FLJ14451 protein with PML-NBs would shed some light on its function since PML-NBs contain mainly proteins involved in cell cycle control and apoptosis such as PML, p53, SUMO and DAXX (Bernardi and Pandolfi 2003). To investigate whether the FLJ14451 colocalizes with the PML-NB specific proteins SP100 and PML-1, MCF-7 cells were transiently cotransfected with expression vectors for FLJ14451 and GFP-fusions of SP100 or PML-1. Figure 21 clearly shows that FLJ14451 does not localize to PML bodies.

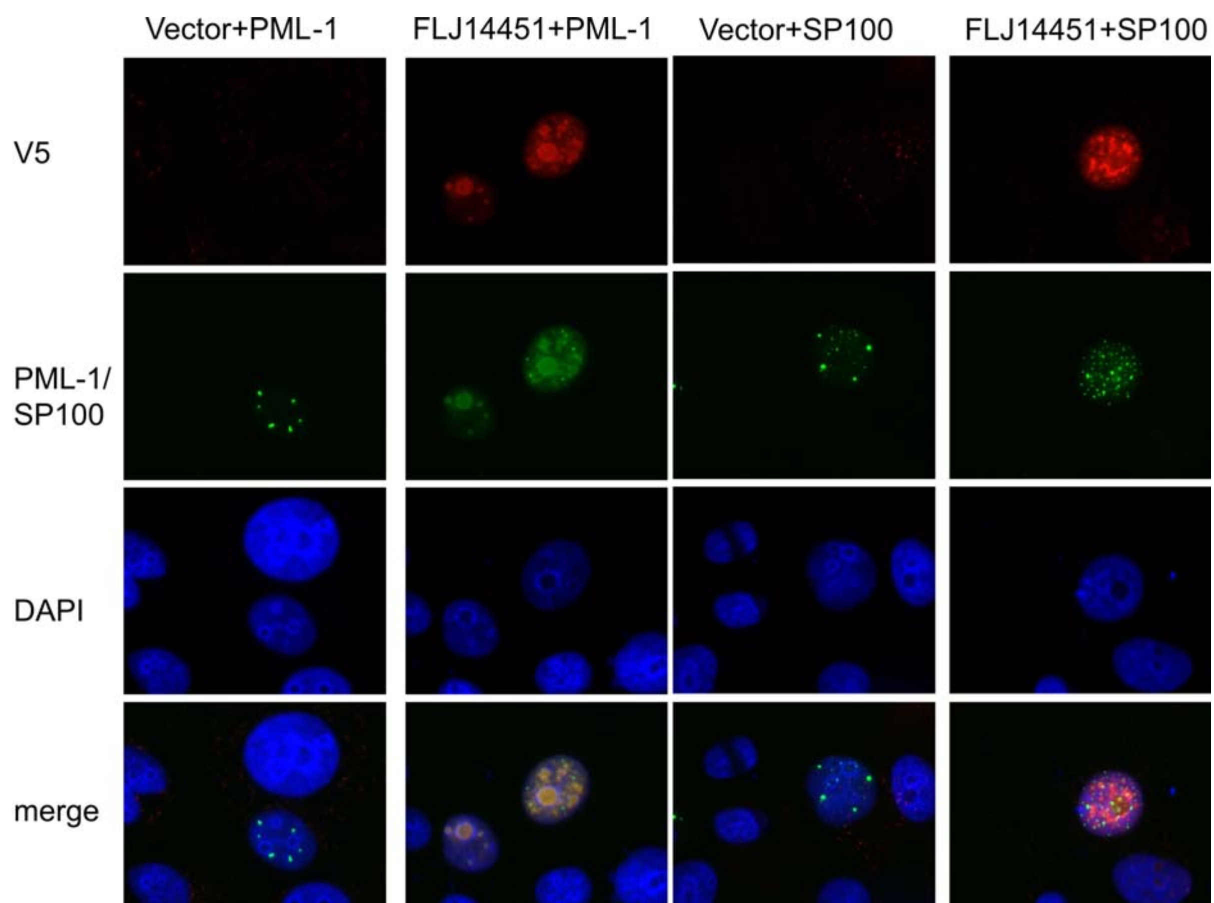


Figure 21. PML-NB and FLJ14451 is not concentrated in PML nuclear bodies. MCF-7 cells were transiently cotransfected with expression vectors encoding the FLJ14451-V5 fusion and a SP100gfp or PML1gfp fusion. Control transfections were performed with an empty V5-vector. The FLJ14451 protein was detected by immunofluorescence as before (red), SP100 and PML-1 via the gfp moiety (green); DAPI was used to counterstain nuclei (blue).

4.3.5. Over-expression of FLJ14451 protein inhibits colony formation in soft agar

To directly test whether the selected candidate genes are linked to some sort of oncogenic process, MCF-7 clones stably over-expressing FLJ14451 and ZFP67 were isolated and tested for anchorage independent growth in semisolid cultures (Grossmann 2002; Wang 2004). As shown in figure 22, *FLJ14451* but not *ZFP67* significantly inhibited MCF-7 colony formation in soft agar cultures. Since this inhibition clearly correlated with protein expression, the results suggest that *FLJ14451* might have tumor suppressor functions.

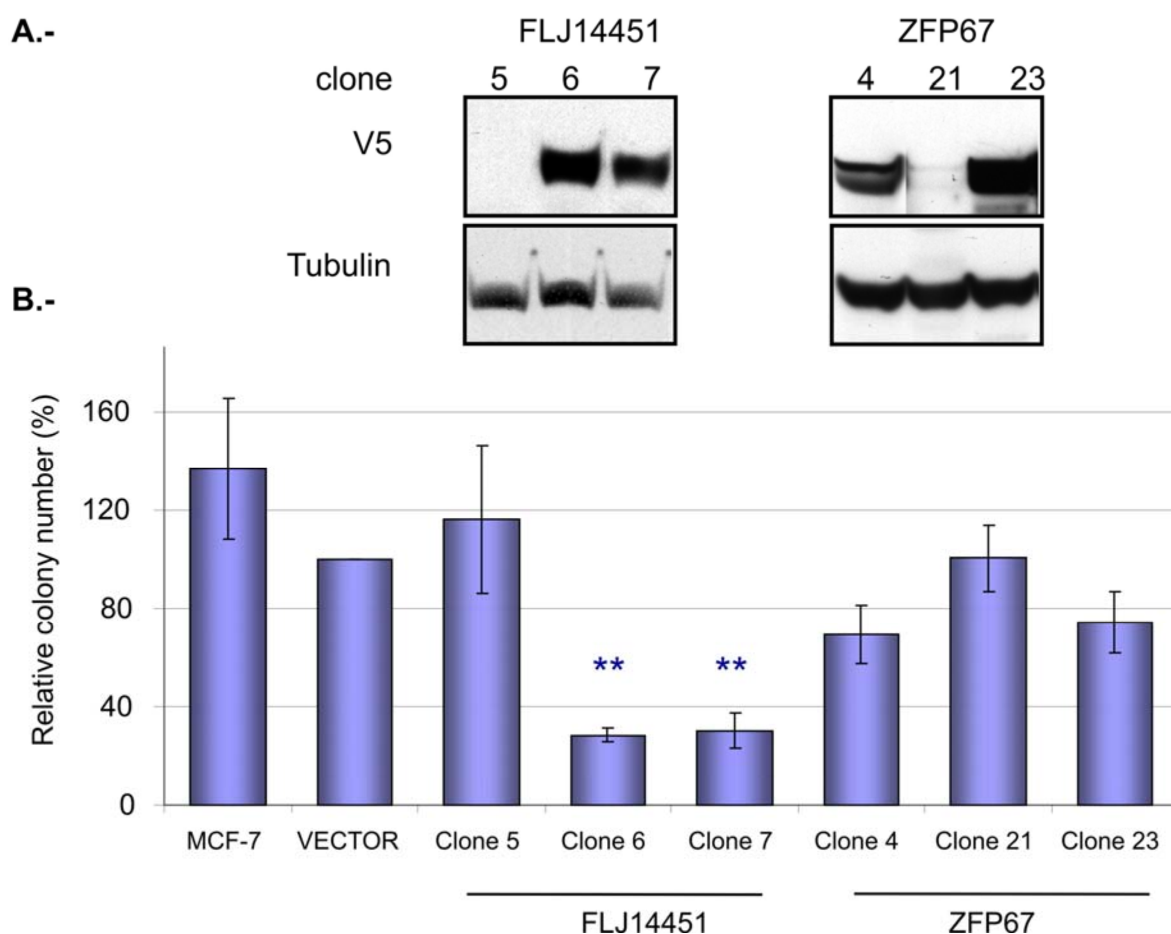


Figure 22. FLJ14451 suppresses colony formation in soft agar. A pool of MCF-7 cells stably transfected with an empty expression vector (vector) and six different clones stably transfected with an FLJ14451-V5 or ZFP67-V5 expression vector were assayed for colony formation in soft agar. A.- Expression of fusion proteins. Lysates (90 µg) of the clones used for the experiment were analyzed in a Western blot using an anti-V5 antibody; tubulin was used as a loading control. B.- Colony formation in soft agar. 500 cells in 1 ml soft agar were plated into the wells of a 6 well plate containing 2 ml bottom agar. For each cell line duplicate plating was performed. After 21 days at 37°C colonies were counted under a microscope. The bar graph shows the colony number relative to the cell population stably transfected with the empty expression vector. Values represent the mean and standard error of three independent experiments (** ANOVA $p<0.001$).

5. - Discussion

5.1. Recovery of TNFα induced genes by combined gene trap mutagenesis and site specific recombination

By combining gene trap mutagenesis with site specific recombination a strategy was developed which enriches for genes induced by TNFα in the human breast cancer cell line MCF-7 even if the genes are transiently or weakly induced. The strategy relies on a one way gene expression switch which converts the transient activation of a gene trap encoded Cre recombinase into the stable expression of a selectable marker. Two selection steps enrich for gene trap integrations into genes specifically activated by the stimuli (here TNFα).

In contrast to conventional gene trapping the present strategy allows screening of a gene trap event that responded to a signaling molecule, as described in the introduction, the Cre induced irreversible recombination in the gene expression switch, allows the gene trap expression to be temporally abrogated. Moreover, this Cre/*loxP* system is very sensitive and can thus identify genes, which are expressed at low levels (Thorey, Muth et al. 1998). This sensitivity is also reflected in a similar study using the ROSACre/*loxP* system (reverse orientation splice acceptor Cre) in mouse embryonic stem (ES) cells. The proportion of active gene traps detected in these experiments was 42.6%, which is much higher than the 11.6% that reported with the standard gene trap vector ROSAβ-gal (Chen, Liu et al. 2004).

In designing the Cre/*loxP* strategy for the recovery of genes induced by TNFα, the apoptotic effects of the TNFα signaling had to be addressed. Binding of TNFα to its receptor TNFR1 recruits the adapter protein TNFR-associated death domain (TRADD). This activated receptor then serves as an assembly platform for binding of several molecules which constitute the first complex that stimulates survival pathways through the activation of NF-κB (canonical pathway) and JNK (see introduction section 4.1.). A second complex (complex II), the death-inducing signaling complex (DISC) which lacks TNFR1 but includes Fas-associated death domain (FADD) and pro-caspases-8 and -10 is subsequently formed in the cytoplasm by modification of complex I. This secondary complex initiates the apoptotic response by activation of the pro-caspases (Barnhart and Peter 2003; Micheau and Tschopp 2003). The apoptotic pathway can be circumvented by over-expression of a N-terminally truncated FADD protein (dnFADD) which has been shown to exert a dominant negative effect on

receptor mediated apoptosis including that induced by TNFα. Therefore, a cDNA encoding dnFADD was used in this project as a second marker in the reporter construct.

With regard to TNFα regulation of target genes it is important to mention that TNFα can induce two modes of NF-κB activation patterns. In the monophasic mode, which is the result of a brief TNFα stimulus, NF-κB enters the nucleus and induces the expression of IκB proteins whose synthesis redistributes NF-κB into the cytoplasm restoring cellular homeostasis. In contrast, persistent TNFα activation produces prolonged NF-κB activation and continued IκB proteolysis, resulting in repeated rounds of NF-κB translocation and cytoplasmic recapture. This activation profile is characterized by a series of asynchronous damped oscillations of nuclear NF-κB (Nelson, See et al. 2004). Stimulation experiments producing monophasic or oscillatory modes have shown that the oscillatory mode is required for late gene expression (Tian, Nowak et al. 2005).

Aside NF-κB activation, stimulation of c-Jun N-terminal kinase, which activates the transcription factor complex AP-1, is a second cellular response to TNFα common to all cell types (see introduction section 4.1.). Therefore, the recovered genes could also be targets of AP-1. In that context it is interesting to note that Banno and colleagues indicated that most of the cell cycle regulators, RNA-processing and metabolic enzymes induced by TNFα are not NF-κB dependent (Banno, Gazel et al. 2005). In depth promoter analyses would be necessary to reveal the dependence of the recovered genes on NF-κB and/or AP-1.

Due to the sensitivity of the Cre/*loxP* system only 21 insertions were in protein coding genes of which some are known to be regulated by TNFα, thus validating the system; these proteins include:

(i) JunB, a key transcriptional regulator of myelopoiesis and a potential tumor suppressor (Passegue et al., 2001; Schwamborn et al., 2003; Tian et al., 2005b; Zhou et al., 2003).

(ii) SLC1A2 (GLT-1/EAAT2), which is responsible for the clearance of the neurotransmitter glutamate from neuronal synapses in the central nervous system (CNS) and is down-regulated by TNFα in a NF-κB dependent manner. Impaired glutamate uptake by glial cells induces excitotoxic neuronal death as a result of glutamate receptor over-stimulation. As a result, mice lacking SLC1A2 develop progressive neuro-degeneration and epilepsy. TNFα mediated

inhibition of SLC12A2 by NF-κB may contribute to glutamate toxicity and cell death in neuro-inflammation and disease (Sitcheran, Gupta et al. 2005).

(iii) ABCC3, one of the hepatocellular ATP-binding cassette proteins transporting bile salts, 17β estradiol and anti-cancer drugs (like doxorubicin or etoposide) across both intracellular and extracellular membranes. Indeed, the acquisition of drug resistance in cancer cells is often associated with increased expression of various cell surface ABC transporters (Tada, Wada et al. 2002). Induction of ABBC3 is hepato-protective in cholestasis and is dependent on intact TNFα signaling pathways (Bohan, Chen et al. 2003).

Several other strategies have been used to identify TNFα regulated genes, among them microarray analysis (Schwamborn, Lindecke et al. 2003; Zhou, Scoggin et al. 2003; Banno, Gazel et al. 2005; Thiefes, Wolter et al. 2005; Tian, Nowak et al. 2005). Numerous microarray studies have been published involving various cells types and TNFα treatment protocols; as a result the shown to be regulated in the different surveys, were quite heterogeneous. Thus, it is not surprising that only a few genes recovered by expression profiling were also identified by gene trapping. In addition, microarray analyses report steady state RNA levels at fixed time points, whereas the gene trap approach used here relies on transcriptional induction within a time window, which if transient could escape detection by the chip based method. This might provide another explanation for the observed differences between the two techniques.

Overall, significantly fewer TNFα regulated genes were recovered by gene trapping than by the microchip approach. This is mainly because cells harboring gene trap integrations in TNFα inducible genes already expressed in the non-induced state were eliminated during the initial selection, which might also explain why only a few previously characterized TNFα inducible genes were recovered.

Microarray studies have indicated that TNFα mediated gene repression is far less frequent than induction. For example, Zhou and colleagues, who performed a microarray analysis in HeLa cells treated with TNFα, found no genes that were down-regulated more than 2 fold (Zhou, Scoggin et al. 2003). In contrast, analysis of U373 human glioblastoma cells treated with TNFα recovered down-regulated genes, but also indicated that down-regulation by a factor of two or more was rarer than up-regulation (Schwamborn, Lindecke et al. 2003). In the

present study only three genes were significantly down-regulated by TNFα: *S100A10*, *SOM/TFCP2L4* and *FGD3*. The recovery of genes down-regulated in the present study is related to two factors: gene trap integration in *FGD3* and in *S100A10* is on the non-coding strand of these genes indicating that most likely the trapped RNA is not the *FGD3* or *S100A10* mRNA, but an antisense transcript from the same genomic region. In contrast, the *SOM*/*TFCP2L4* gene trap integration is on the coding strand suggesting up-regulation of this gene at other time points which is shown in the northern blot analysis.

Based on their response to TNFα in wild type cells the trapped genes belonged to one of the following four categories: (i) Non-responsive genes. Although no transcriptional regulation was observed in this category it is possible that induction does occur at some time point not covered by the Northern blots; (ii) genes that were either permanently induced or repressed, and (iii) genes that were only transiently induced.

Interestingly, not only regulation itself, but also the specific time points at which it became detectable differ between the genes. Some genes such as *BCL9L*, *FGD3*, *JunB* and *SOM/TFCP2L4* responded to TNFα already after 4 hours. Others such as *ABCC3*, *C1QTNF6*, *S100A10*, *SLC12A2* and *WDR10* required up to 24 hours to respond. Slow responding genes are likely to be downstream targets of transcription factors induced early in the process. In line with this, some of the recovered genes indeed code for transcription factors (*ESR1*, *JunB*, *SOM/TFCP2L4*, *ZFP67*, *ZNF143*), which can activate or repress specific target genes.

These different patterns, transient and late/early induction, may reflect the complex kinetics of recruitment of NF-κB to its targets. These kinetics have been studied by CHIP analysis and led to the classification into two groups of genes (Saccani, Pantano et al. 2001). One subset of target genes was found to be occupied with heavily acetylated histones already before stimulation and therefore was subsequently accessible immediately after NF-κB activation. The mechanism of rapid transcriptional induction has also been associated with the pre-formation of a transcription competent initiation complex on the promoters (Ainbinder, Revach et al. 2002). In contrast, other target genes had low histone acetylation levels and became occupied only 90-120 min after NF-κB nuclear entry (Saccani, Pantano et al. 2001).

5.2. Mechanisms of entrapment

A large number of the gene trap insertions, like the ones located upstream of genes or in introns or in opposite orientation to annotated transcripts, cannot simply be explained, as U3Cre is conceived as an exon trap. The exon trap is completely devoid of regulatory sequences and therefore Cre transcription is dependent on the expression of the trapped gene. Because of the viral preference the majority of integrations will be at the 5'-end of genes. The integrated U3Cre then functions as a 3'-terminal exon with transcription terminating at border of the proviral R and U5 sequence elements.

Interestingly, similar gene traps used for high throughput mutagenesis in ES cells also showed that only 20% of all the insertions present in the library were in exons, whereas the majority was in introns.

The U3Cre gene trap inserts a partial terminal 3'-exon, which has a polyA site and no splice acceptor. Generation of an unspliced fusion transcript between cellular sequences and the gene trap encoding Cre should be a rare event, because polyadenylation and 3'-end formation are inefficient when polyA sites are placed between a 5'-splice donor and a 3'-splice acceptor site as in an intron (Adami and Nevins 1988; Osipovich, White-Grindley et al. 2004).

Alternatively, the U3Cre gene trap could have activated cryptic splice acceptors 5' to the integration site. As a consequence a fusion transcript would be produced, that close to the Cre containing exon, contains more or less intronic sequence, depending on the position of the splice acceptor. These cryptic splice acceptors could be localized within the intron, the provirus sequence or might be generated by the virus insertion.

Analysis of one such case from the work described here, where the U3Cre gene trap had integrated into the first intron of the *TFCP2L4* gene, revealed a fusion transcript initiating in the cellular exon 1 and splicing into the proviral sequence downstream of the 5'-LTR (J. Altschmied, C. Strolz, personal communication). This kind of splicing might also occur in the remainder of the sense integrations in the first intron (8 out of 21 sense integrations). Nevertheless, transcriptional activation of the gene trap encoded recombinase has to be regulated in order to be positively selected. To explain this type of integrations it has to be assumed that before TNFα induction there was no Cre encoding transcript present, which was the case for the above mentioned integration in *TFCP2L4*.

Due to the sensitivity of the present system these integration events could involve other different processes in addition to the cryptic splicing such as trapping of additional, so far unknown transcripts, or the activation of alternative or bidirectional promoters.

Alternative promoters have been already described in the human genome (Kim, Barrera et al. 2005) and constitute prime target elements through which diversity and flexibility are created (Ayoubi and Van De Ven 1996). A recent report from the RIKEN, FANTOM and Genome Science groups assigned an average of 1.32 5'-start sites for each 3'-end (Carninci, Kasukawa et al. 2005), which indicates a widespread presence of alternative promoters.

Many of the genes, which have been reported to have multiple promoters, show no variation in the primary structure of the resulting protein, however, the mRNA variants differ in their transcriptional patterns and translational efficiencies. In these cases alternative promoter usage results in variant 5'-UTRs that might differ by the presence of an upstream ORF which can affect translation (reviewed by Landry, Mager et al. 2003). Well known examples of genes with alternative promoters, whose products have an influence on apoptosis or proliferation are *p53*, *p21*, or *c-myc*.

An example for alternative promoter usage in the present study is *TFCP2L4/SOM*, a gene with known multiple transcripts, where U3Cre had integrated into the first intron. The use of alternative first exons and differential splicing of exon 2, results in three protein isoforms (Ting, Wilanowski et al. 2003). Preliminary experiments indicate that two of these protein variants can induce cell migration in human umbilical vein endothelial cells, (J. Altschmied, C.Strolz, J.Haendeler, C. Schön, personal communication).

Lastly, another possibility is that these integrations could represent additional, distinct transcripts. The U3Cre gene trap "selects" for events where Cre provides the first AUG in the resulting transcript (von Melchner, Reddy et al. 1990), therefore intron or exon integrations in regions distant from the UTR are expected not to provide the first AUG for translation. However there are several examples where U3Cre is integrated far downstream of the first exon. This is the case for *ABCC3*, *SLC12A2* and *SPATA 5* where the gene trap is integrated in intron 14, 8 and 5, respectively. These integrations could report the presence of uncharacterized transcriptional units embedded in the same orientation in the apparent host gene. Though this possibility appears unlikely, it is not impossible, as there are examples of

such a situation in the human genome, in which the embedded transcriptional unit consist of one unique exon (Bejanin, Cervini et al. 1994; Conrad, Vianna et al. 2002).

Antisense U3Cre integrations: bidirectional promoters or antisense transcripts?

A striking point in the present study is that a large proportion of the gene trap integrations are on the non-coding strand of known transcripts (see results section 4.2.3.). Antisense U3Cre integrations represent roughly 40% of all integrations in annotated genes and Genscan gene predictions. The high percentage of integrations on the non-coding strand of transcripts might be due to the stringent selection for integrations in genes specifically regulated by TNFα and the MMLV preference for integrations into actively transcribed regions. This combination would enrich for integration events into genes not active in the absence of TNFα, which are at the same time located in actively transcribed regions.

One of the most intriguing questions regarding the antisense gene trap integrations is the nature of the recovered transcripts. A recent report indicates that mouse natural sense-antisense transcripts tend to be polyA negative and nuclear localized (Kiyosawa, Mise et al. 2005); this and other options will be discussed in detail in section 5.3.

Independent of the nature of the trapped RNA is the mechanism gene trap activation, when the trap is antisense relative to an annotated gene. In particular, integrations close to the 5'-end could reflect the activity of a bidirectional promoter or could be indicative for the presence of a long transcript originating within the gene, a so called naturally occurring antisense transcript (NAT).

Promoters are the central processors of transcriptional control. They comprise the genomic DNA sequences found upstream of transcription units and sometimes extend into the first exon of a gene. Interestingly, analysis of the human genome indicated that transcript pairs arranged head to head with less than 1000 bp separating their transcriptional start sites are controlled by bidirectional promoters. A common feature of these bidirectional promoters is the presence of a CpG island between the genes which can overlap partially or entirely with the first exons (Adachi and Lieber 2002; Trinklein, Aldred et al. 2004).

Individual examples of bidirectional gene pairs have been known for years, some serve to maintain a stoichiometric relationship between proteins required in a certain ratio (histones,

Ahn and Gruen 1999), others regulate the co-expression of genes that function in the same pathway (collagen type IV genes, Schmidt, Fischer et al. 1993), control transcripts in a temporal fashion (genes regulated during the cell cycle Guarguaglini, Battistoni et al. 1997) or provide coordinated responses to induction signals such as in the case of heat shock protein genes (Hansen, Bross et al. 2003).

A comparison of the activity and directionality of 258 bidirectional promoters with 56 random promoters in four cell lines revealed that the cellular context is important for the bidirectional activity of a promoter. Whereas 33% of the random promoters analyzed showed bidirectional activity in half of the cell lines and unidirectional activity in the other half, 29% showed bidirectional activity in all four cell lines. In contrast, 57% of the bidirectional promoters assayed showed bidirectional activity in all four cell lines (Trinklein, Aldred et al. 2004). Consistent with this study, two non exclusive scenarios would be possible to explain U3Cre integrations found 5'-upstream and in antisense orientation to known transcripts. The orientation of the provirus could reflect the presence of a bidirectional promoter with stringent transcriptional activation in the sense and/or antisense direction in a specific cellular context. Alternatively, such integrations could reflect the presence of transcriptional noise coming form a spurious bidirectional promoter, in this case the transcript coming from the opposite strand might be stabilized by the presence of a complete reading frame and a polyA tail. In the present project U3Cre integrations in genomic regions with transcripts arranged head to head with less than 1000 bp that might represent bidirectional transcript pairs, were observed in 7 integrations events. In two of these cases both transcripts are derived from annotated genes, whereas in the others the pairs consist of a fully annotated transcript on one strand and an EST or Genscan prediction on the other.

On the other hand, U3Cre integrations on the non-coding strand could report natural antisense transcripts. In recent years NATs have been linked to many aspects of eukaryotic gene expression including genomic imprinting, RNA interference, X-chromosome inactivation and RNA editing. NATs have also been implicated in some diseases, an example being a heritable α-thalassemia, where a chromosomal deletion results in juxtaposition of a truncated, widely expressed gene (*LUC7L*) close to a structurally normal α globin gene (*HBA2*). As a result an antisense transcript originating from the *LUC7L* promoter extends into the *HBA2* CpG island leading to methylation and silencing of the gene (Kleinjan and van Heyningen 2003;

Tufarelli, Stanley et al. 2003). Several computational approaches have been developed to identify antisense transcripts (Li et al., 2006; Shendure and Church, 2002; Yelin et al., 2003) and have supported the view that antisense regulation might be more pervasive in the genome than previously appreciated.

The current strategy yielded only four U3Cre integrations, in which the gene trap had integrated close to regions, where two annotated, overlapping RNAs are transcribed from the two DNA strands. In three cases the complete transcripts on both strands are known (*RACK7/Q8TE85*, *S100PBP/YARS*, *TFCP2L4/OTTHUMG00000003039*), whereas in the other an EST is found on the non-coding strand within an intron of a known gene (*TFF1*). However, so far nothing is known about the function of these antisense transcripts or their regulation under specific conditions. The overlapping region between the transcripts can be either in introns or exons, the later is represented only by one pair (*S100PBP/YARS*), where both transcripts overlap with their respective UTRs.

Noteworthy, in spite of not being within transcripts in the same transcriptional direction, some of the gene trap integrations on the non-coding strand of known or hypothetical protein genes are downstream of putative transcription start sites, which in most instances are located in the first intron or exon of the non-coding strand of these genes.

It would not be surprising if the antisense transcripts were implicated in regulatory processes after TNFα treatment, as there have been examples for regulation through NATs. A natural collagenα1(I) antisense transcript is found in chicken chondrocytes, where the sense transcript level is low. Upon 5-bromo-2'-deoxyuridine (BrdU) treatment, antisense transcription decreases while expression of the sense transcript rises, being exactly correlated with overall collagenα1(I) mRNA accumulation. These results suggest that either the activity of the sense and antisense promoters is differentially regulated and/or that there is interference between sense and antisense transcription (Farrell and Lukens 1995). Other examples of regulation in opposite direction are the regulation of the thymidine kinase (Sutterluety, Bartl et al. 1998) and eIF-2α (Silverman, Noguchi et al. 1992; Noguchi, Miyamoto et al. 1994) genes, where antisense promoters have been described in intron 3 and intron 1, respectively.

Following the example of collagenα1(I), the expression of genes with U3Cre integrations on the non-coding strand would be expected to be down-regulated as result from the transcription

of the non-coding strand. Surprisingly, in the present work, northern blot analysis of candidate genes in most cases showed no correlation between gene trap orientation and the regulation of the coding transcript. This absence of a correlation between sense/antisense pairs has also been observed in a study from the RIKEN, FANTOM and Genome Science groups, which showed that most of the sense/antisense gene pairs were positively correlated in their expression. A possible explanation for this co-expression would be that the transcription of the sense/antisense pairs is controlled by the same enhancer elements (Katayama, Tomaru et al. 2005).

In some instances, in particular gene trap integrations antisense and upstream to known genes, a transcript involved in the regulation of an alternative promoter might have been trapped. This could be the case for the integrations in the genes coding for topoisomerase 1, ckrox or the estrogen receptor 1. For the first there is an exon prediction upstream to the experimentally validated first exon, the upstream regions of the latter two genes are characterized by the presence of several alternative exons. All these examples may have alternative promoters active under different cellular contexts, where antisense transcripts could help to fine-tune the transcript levels of hypothetical alternative variants in response to TNFα.

5.3. Nature of the recovered transcripts

Polyadenylated versus non-polyadenylated

As far as the nature of the trapped transcripts is concerned, these do not necessarily require to be polyadenylated. Since polyadenylation stabilizes transcripts, polyA negative transcripts are harder to detect presumably due to a shorter half life. By providing a polyA tail, U3Cre insertions presumably stabilize such transcripts making them easier to detect.

Transcriptome analyses have traditionally focused on cytoplasmic polyA RNA to exclude rRNAs, tRNAs and incompletely processed primary transcripts. It was then assumed that most transcripts were derived from protein coding genes, processed to polyadenylated mRNAs, which were transported to the cytoplasm for translation (Frith, Pheasant et al. 2005).

Histone RNAs have been long considered to be the only transcripts synthesized without polyA tails (Birnstiel, Busslinger et al. 1985). Moreover, 30% of the RNAs associating with

polysomes in actinomycin D-treated HeLa cells were found to be polyA negative (Milcarek, Price et al. 1974) and in Chinese hamster cells the fraction of polysome associated transcripts with 5'-cap structures exceeds the polyA positive fraction by 3 fold (Salditt-Georgieff, Harpold et al. 1981). Recent analyses of the transcriptome indicated that the number of non-polyA transcripts is higher than expected and that polyA negative RNAs constitute the bulk of the RNA in both nucleus and cytoplasm (Cheng, Kapranov et al. 2005).

New transcripts

Two large scale efforts sponsored independently by the National Cancer Institute (Strausberg, Feingold et al. 1999; Strausberg, Feingold et al. 2002) and RIKEN (Okazaki, Kikuno et al. 2004; Ota, Suzuki et al. 2004) are based on the identification of new cDNAs. However, many transcripts will escape this type of analyses, because they are present in small quantities or only transiently. With the present strategy some of these transcripts might have been localized.

As might have been expected from the sensitivity of the gene trap, bioinformatics analysis of gene trap sequence taqs (GTSTs) has shown that some of the U3Cre integrations were in unannotated genomic regions, which in several instances contained putative transcriptional start sites or predicted transcripts. These transcriptional start sites were localized several hundred or thousand bp upstream of the gene trap integrations, supporting the presence of a transcript within these regions. Integrations into such regions could reflect the specific activation of transcripts in a defined cellular context (TNFα stimulus). If such transcripts can be validated experimentally, this would indicate that gene trapping with such a system of high sensitivity could be helpful in the annotation of the human genome sequence.

Coding versus non-coding

Recently, genome tiling arrays (Frith, Pheasant et al. 2005) have shown that the transcribed fraction of the human genome is not limited to protein coding genes and is much larger than previously thought including introns and many not yet annotated regions. Part of the non-coding transcripts could simply result from a low level transcriptional noise without having any biological significance. Alternatively, these transcripts may help to increase accessibility

of nearby protein coding genes for regulatory proteins (Cheng, Kapranov et al. 2005; Kapranov, Drenkow et al. 2005). About 98% of all transcriptional output mapped to the human genome is non-coding, including the untranslated regions and introns of protein coding RNAs (Mattick and Makunin 2005). Actually, only 1.2% of the annotated human genome encodes proteins, although a much larger fraction is transcribed (Frith, Pheasant et al. 2005).

As gene prediction algorithms like Genscan are based on structural features including translational signals, it is likely that integrations in unannotated genomic regions reflect trapping of non-coding RNAs, which could belong to different categories:

(i) Large non-coding RNAs (ncRNAs); such transcripts seem to be involved in many processes. Willingham and colleagues who were able to characterize the biological function of some of these ncRNAs, identified one of them as a modulator of the transcriptional activity of the nuclear factor of activated T cells (NFAT), a function which is achieved by regulating the subcellular localization of the protein. Another one was found to function as a repressor of Hedgehog signaling and six ncRNAs were described to be essential for cell viability (Willingham, Orth et al. 2005). In addition, ncRNAs apparently can contribute to local chromatin modification or methylation when they overlap with promoters.

(ii) Small non-coding double stranded (ds) RNAs. One such dsRNAs has been proven to play a critical role in mediating neuronal differentiation, the mechanism of action appears to be mediated through a dsRNA protein interaction (Kuwabara, Hsieh et al. 2004).

(iii) Small non-coding RNAs. Recently transcripts, which are derived from the thritorax response elements (TREs) located in the *ultrabithorax* (*Ubx*) locus, have been shown to play an important role in the activation of gene expression. These ncRNAs transcribed through the TREs are retained at the TREs by DNA-RNA interactions and provide a RNA scaffold that is bound by Ash1, a protein without intrinsic DNA-binding capabilities, which is essential for the expression of the homeotic gene *Ubx* (Sanchez-Elsner, Gou et al. 2006).

(iv) Another class of non-coding RNAs, which have received a lot of attention recently, are microRNAs (miRNAs). These are small RNAs processed from pre-microRNAs in the nucleus by the ribonuclease III Drosha. Processing results in small double stranded RNAs with a hairpin structure and unpaired nucleotides at both ends which are recognized by importin 5.

After export to the cytoplasm they are processed by the ribonuclease III Dicer, separated by a helicase and one strand is incorporated into the RNA induced silencing complex (RISC) (Zamore and Haley 2005). RISC can target protein-coding messenger RNA (mRNAs) resulting in a translational block or mRNA degradation. Base pairing between the miRNA and its complementary target mRNA gives the process its specificity. The choice between translation inhibition and destruction is thought to be governed by the degree of mismatch between the miRNA and its target mRNA, with degradation being the outcome for best-matched targets (Meltzer 2005).

Although nowadays a large number of miRNAs are known (Griffiths-Jones 2004; Mattick and Makunin 2005) only a few miRNA targets have been identified in mammals. The physiological importance of miRNAs has been experimentally demonstrated *in vivo* by Krützfeldt and colleagues, who made use of a molecule complementary to miRNA-122 (miR-122) called antagomir-122. Intravenous administration of antagomir-122 in mice resulted in a marked reduction of miRNA-122 levels. Several hundreds of genes were affected by the antagomir. Interestingly, the 3'-UTRs of up-regulated genes were strongly enriched in miRNA-122 recognition motifs whereas down-regulated genes were depleted in these motifs. The functional annotation of the down-regulated genes predicted that cholesterol biosynthesis should be affected. Indeed, plasma cholesterol levels were reduced in antagomir treated animals (Krutzfeldt, Rajewsky et al. 2005).

The present gene trapping strategy has not recovered known human miRNAs genes; however one gene trap integration was localized 10 kbp downstream of a known miRNA (mo-mir21). miRNAs are synthesized as pre-mRNAs which are processed post-transcriptionally, therefore it is conceivable that the gene trap is localized somewhere in a large precursor. Alternatively, the integration could have occurred in a so far unannotated downstream miRNA gene within a miRNA cluster. As a matter of fact, miRNA clustering is significantly higher than expected at random; from the human known miRNA genes analyzed by Altuvia and colleagues 37% appeared in clusters of two or more with pairwise chromosomal distances of less than 3 kbp (Altuvia, Landgraf et al. 2005). Another study suggested that clustered miRNAs are expected to be found within a range of 50 kbp (Baskerville and Bartel 2005).

5.4. Functional validation of the trapped genes

Functional validation was performed with two trapped genes (see results section 4.3.2.); the reasons for selecting these genes were based on the response to TNFα, their availability as full length cDNAs from the RZPD and a reported association with cancer:

a) *ZFP67/c-Krox* belongs to a family of developmentally regulated genes and was shown to repress the expression of extracellular matrix proteins, particularly fibronectin (Widom, Culic et al. 1997). Inhibition of fibronectin in several *in vitro* transformation assays increased the rate of transformation, suggesting that the protein has a tumor suppressor function (Steel and Harris 1989).

b) FLJ14451 is a hypothetical protein whose only obvious feature is the presence of 3 copies of a C-x8-C-x5-C-x3-H type zinc finger domain, which is also found in a class of eukaryotic zinc finger proteins involved in cell cycle regulation.

The evaluation of the influence of the FLJ14451 and ZFP67 proteins on anchorage independent growth showed differences between the two proteins. Over-expression of FLJ14451 in MCF-7 cells suppressed soft agar colony formation, whereas ZFP67 had no effect on this process.

The lack of difference in colony formation in soft agar between *ZFP67* overexpressing clones and control cells is surprising as the ZFP67 protein (cKrox) regulates extracellular matrix genes, which are known to affect oncogenic transformation. One explanation for this discrepancy could be the molecular alterations already present in the MCF-7 cell line which might compensate ZFP67 up-regulation. Moreover, the levels of cKrox could be already so high in the cells that it could saturate all the genomic binding sites and therefore no additional effect would be observed upon ZFP67 up-regulation. Additional down-regulation experiments are required to clarify the role of ZFP67 in oncogenic transformation.

Inhibition of anchorage independent growth by the FLJ14451 protein might be important for controlling the spread of cells from the site of a primary tumor to distant locations. FLJ14451 may be involved in regulation of pathways implicated in cell proliferation and in circumstances where the survival signals fail, like in anchorage independent growth, slow down cell growth or induce apoptosis.

It is well known that adhesion of cells to the extracellular matrix stimulates signal transduction cascades that have been shown to impinge on cell growth, differentiation and cell death. Indeed, detachment of a cell from is supportive matrix induces anoikis, a specific type of apoptosis (Reddig and Juliano 2005), which might be triggered by FLJ14451.

Interestingly, microarray studies support a role of FLJ14451 in maintaining a non-malignant state, specifically in two particular instances. In the first case, FLJ14451 showed progressive down-regulation in breast tumors from grade 1 to 3 (van 't Veer, Dai et al. 2002). The histological grade of breast carcinomas has long provided clinically important prognostic information. Grade 1 is the lowest grade with well-differentiated cells, grade 2 is the intermediate with moderately differentiated cells and grade 3 is the highest with poorly differentiated and fast growing cells. In the second case, a decrease in the FLJ14451 mRNA levels of primary cancers in patients who developed metastasis after 5 years was reported (van 't Veer, Dai et al. 2002). Altogether these data indicate that FLJ14451 expression could be inversely correlated with malignancy and metastasis. Therefore, the protein might be a new tumor suppressor.

5.5. Future perspectives

Several proteins identified in this study merit future consideration. One of them, *FLJ14451*, is a putative tumor suppressor which needs further validation in both *in vitro* and *in vivo* experiments. The other is the *SOM/TFCP2L4* transcription factor which appears to be a potent inducer of cell migration and is thus likely involved in tumor metastasis and angiogenesis.

The high percentage of gene trap integrations into non-coding regions of the genome suggests that the strategy might be useful in disrupting non-coding genes. Pending further validation of the trapped non-coding transcripts, the combination of gene trap mutagenesis and site specific recombination as described here could be applicable to the high throughput mutagenesis of non-coding genes for which no alternative method is presently available.

6. - References

Adachi, N., and Lieber, M. R. (2002). Bidirectional gene organization: a common architectural feature of the human genome. Cell *109*, 807-809.

Adami, G., and Nevins, J. R. (1988). Splice site selection dominates over poly(A) site choice in RNA production from complex adenovirus transcription units. Embo J *7*, 2107-2116.

Adams, D. J., da Silva, M. W., Flowers, J. L., Kohlhagen, G., Pommier, Y., Colvin, O. M., Manikumar, G., and Wani, M. C. (2006). Camptothecin analogs with enhanced activity against human breast cancer cells. I. Correlation of potency with lipophilicity and persistence in the cleavage complex. Cancer Chemother Pharmacol *57*, 135-144.

Aggarwal, B. B., Kumar, A., and Bharti, A. C. (2003). Anticancer potential of curcumin: preclinical and clinical studies. Anticancer Res *23*, 363-398.

Ahn, J., and Gruen, J. R. (1999). The genomic organization of the histone clusters on human 6p21.3. Mamm Genome *10*, 768-770.

Ainbinder, E., Revach, M., Wolstein, O., Moshonov, S., Diamant, N., and Dikstein, R. (2002). Mechanism of rapid transcriptional induction of tumor necrosis factor alpha-responsive genes by NF-kappaB. Mol Cell Biol *22*, 6354-6362.

Ali, S., and Coombes, R. C. (2000). Estrogen receptor alpha in human breast cancer: occurrence and significance. J Mammary Gland Biol Neoplasia *5*, 271-281.

Altuvia, Y., Landgraf, P., Lithwick, G., Elefant, N., Pfeffer, S., Aravin, A., Brownstein, M. J., Tuschl, T., and Margalit, H. (2005). Clustering and conservation patterns of human microRNAs. Nucleic Acids Res *33*, 2697-2706.

Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, I.G., Smith, J.A. und Struhl, K. (1989) *Current protocols in molecular biology.*, New York, Chichester, Brisbane, Toronto, Singapore

Ayoubi, T. A., and Van De Ven, W. J. (1996). Regulation of gene expression by alternative promoters. Faseb J *10*, 453-460.

Bader, S., Walker, M., and Harrison, D. (2000). Most microsatellite unstable sporadic colorectal carcinomas carry MBD4 mutations. Br J Cancer *83*, 1646-1649.

Baer, A., Schubeler, D., and Bode, J. (2000). Transcriptional properties of genomic transgene integration sites marked by electroporation or retroviral infection. Biochemistry *39*, 7041-7049.

Bajic, V. B., and Seah, S. H. (2003). Dragon Gene Start Finder identifies approximate locations of the 5' ends of genes. Nucleic Acids Res *31*, 3560-3563.

Bajic, V. B., Seah, S. H., Chong, A., Zhang, G., Koh, J. L., and Brusic, V. (2002). Dragon Promoter Finder: recognition of vertebrate RNA polymerase II promoters. Bioinformatics *18*, 198-199.

Bajic, V. B., Tan, S. L., Suzuki, Y., and Sugano, S. (2004). Promoter prediction analysis on the whole human genome. Nat Biotechnol *22*, 1467-1473.

Balkwill, F. (2002). Tumor necrosis factor or tumor promoting factor? Cytokine Growth Factor Rev *13*, 135-141.

Balkwill, F., and Coussens, L. M. (2004). Cancer: an inflammatory link. Nature *431*, 405-406.

Balkwill, F., and Mantovani, A. (2001). Inflammation and cancer: back to Virchow? Lancet *357*, 539-545.

Banno, T., Gazel, A., and Blumenberg, M. (2005). Pathway-specific profiling identifies the NF-kappa B-dependent tumor necrosis factor alpha-regulated genes in epidermal keratinocytes. J Biol Chem *280*, 18973-18980.

Barnhart, B. C., and Peter, M. E. (2003). The TNF receptor 1: a split personality complex. Cell *114*, 148-150.

Baskerville, S., and Bartel, D. P. (2005). Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. Rna *11*, 241-247.

Bejanin, S., Cervini, R., Mallet, J., and Berrard, S. (1994). A unique gene organization for two cholinergic markers, choline acetyltransferase and a putative vesicular transporter of acetylcholine. J Biol Chem *269*, 21944-21947.

Bernardi, R., and Pandolfi, P. P. (2003). Role of PML and the PML-nuclear body in the control of programmed cell death. Oncogene *22*, 9048-9057.

Bernards, R., and Weinberg, R. A. (2002). A progression puzzle. Nature *418*, 823.

Birnstiel, M. L., Busslinger, M., and Strub, K. (1985). Transcription termination and 3' processing: the end is in site! Cell *41*, 349-359.

Bohan, A., Chen, W. S., Denson, L. A., Held, M. A., and Boyer, J. L. (2003). Tumor necrosis factor alpha-dependent up-regulation of Lrh-1 and Mrp3(Abcc3) reduces liver injury in obstructive cholestasis. J Biol Chem *278*, 36688-36698.

Boris-Lawrie, K. A., and Temin, H. M. (1993). Recent advances in retrovirus vector technology. Curr Opin Genet Dev *3*, 102-109.

Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem *72*, 248-254.

Brembeck, F. H., Schwarz-Romond, T., Bakkers, J., Wilhelm, S., Hammerschmidt, M., and Birchmeier, W. (2004). Essential role of BCL9-2 in the switch between beta-catenin's adhesive and transcriptional functions. Genes Dev *18*, 2225-2230.

Brown, R. T., Ades, I. Z., and Nordan, R. P. (1995). An acute phase response factor/NF-kappa B site downstream of the junB gene that mediates responsiveness to interleukin-6 in a murine plasmacytoma. J Biol Chem *270*, 31129-31135.

Bryson, J. M., Coy, P. E., Gottlob, K., Hay, N., and Robey, R. B. (2002). Increased hexokinase activity, of either ectopic or endogenous origin, protects renal epithelial cells against acute oxidant-induced cell death. J Biol Chem *277*, 11392-11400.

Burge, C., and Karlin, S. (1997). Prediction of complete gene structures in human genomic DNA. J Mol Biol *268*, 78-94.

Burger, M. J., Tebay, M. A., Keith, P. A., Samaratunga, H. M., Clements, J., Lavin, M. F., and Gardiner, R. A. (2002). Expression analysis of delta-catenin and prostate-specific membrane antigen: their potential as diagnostic markers for prostate cancer. Int J Cancer *100*, 228-237.

Burow, M. E., Weldon, C. B., Tang, Y., McLachlan, J. A., and Beckman, B. S. (2001). Oestrogen-mediated suppression of tumour necrosis factor alpha-induced apoptosis in MCF-7 cells: subversion of Bcl-2 by anti-oestrogens. J Steroid Biochem Mol Biol *78*, 409-418.

Burow, M. E., Weldon, C. B., Tang, Y., Navar, G. L., Krajewski, S., Reed, J. C., Hammond, T. G., Clejan, S., and Beckman, B. S. (1998). Differences in susceptibility to tumor necrosis factor alpha-induced apoptosis among MCF-7 breast cancer cell variants. Cancer Res *58*, 4940-4946.

Bushman, F., Lewinski, M., Ciuffi, A., Barr, S., Leipzig, J., Hannenhalli, S., and Hoffmann, C. (2005). Genome-wide analysis of retroviral DNA integration. Nat Rev Microbiol *3*, 848-858.

Campbell, K. J., Rocha, S., and Perkins, N. D. (2004). Active repression of antiapoptotic gene expression by RelA(p65) NF-kappa B. Mol Cell *13*, 853-865.

Campisi, J. (2005). Senescent cells, tumor suppression, and organismal aging: good citizens, bad neighbors. Cell *120*, 513-522.

Carlsson, H., Petersson, S., and Enerback, C. (2005). Cluster analysis of S100 gene expression and genes correlating to psoriasin (S100A7) expression at different stages of breast cancer development. Int J Oncol *27*, 1473-1481.

Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M. C., Maeda, N., Oyama, R., Ravasi, T., Lenhard, B., Wells, C.*, et al.* (2005). The transcriptional landscape of the mammalian genome. Science *309*, 1559-1563.

Cerhan, J. R., Anderson, K. E., Janney, C. A., Vachon, C. M., Witzig, T. E., and Habermann, T. M. (2003). Association of aspirin and other non-steroidal anti-inflammatory drug use with incidence of non-Hodgkin lymphoma. Int J Cancer *106*, 784-788.

Chan, A. T., Giovannucci, E. L., Schernhammer, E. S., Colditz, G. A., Hunter, D. J., Willett, W. C., and Fuchs, C. S. (2004). A prospective study of aspirin use and the risk for colorectal adenoma. Ann Intern Med *140*, 157-166.

Chang, Y. S., Yeh, K. T., Yang, M. Y., Liu, T. C., Lin, S. F., Chan, W. L., and Chang, J. G. (2005). Abnormal expression of JUNB gene in hepatocellular carcinoma. Oncol Rep *13*, 433-438.

Chen, Y. T., Liu, P., and Bradley, A. (2004). Inducible gene trapping with drug-selectable markers and Cre/loxP to identify developmentally regulated genes. Mol Cell Biol *24*, 9930-9941.

Cheng, J., Kapranov, P., Drenkow, J., Dike, S., Brubaker, S., Patel, S., Long, J., Stern, D., Tammana, H., Helt, G.*, et al.* (2005). Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. Science *308*, 1149-1154.

Chinnaiyan, A. M., O'Rourke, K., Tewari, M., and Dixit, V. M. (1995). FADD, a novel death domain-containing protein, interacts with the death domain of Fas and initiates apoptosis. Cell *81*, 505-512.

Cogswell, P. C., Guttridge, D. C., Funkhouser, W. K., and Baldwin, A. S., Jr. (2000). Selective activation of NF-kappa B subunits in human breast cancer: potential roles for NF-kappa B2/p52 and for Bcl-3. Oncogene *19*, 1123-1131.

Conrad, C., Vianna, C., Freeman, M., and Davies, P. (2002). A polymorphic gene nested within an intron of the tau gene: implications for Alzheimer's disease. Proc Natl Acad Sci U S A *99*, 7751-7756.

Davuluri, R. V., Grosse, I., and Zhang, M. Q. (2001). Computational identification of promoters and first exons in the human genome. Nat Genet *29*, 412-417.

De Smaele, E., Zazzeroni, F., Papa, S., Nguyen, D. U., Jin, R., Jones, J., Cong, R., and Franzoso, G. (2001). Induction of gadd45beta by NF-kappaB downregulates pro-apoptotic JNK signalling. Nature *414*, 308-313.

Devarajan, E., Chen, J., Multani, A. S., Pathak, S., Sahin, A. A., and Mehta, K. (2002). Human breast cancer MCF-7 cell line contains inherently drug-resistant subclones with distinct genotypic and phenotypic features. Int J Oncol *20*, 913-920.

Dowen, S. E., Crnogorac-Jurcevic, T., Gangeswaran, R., Hansen, M., Eloranta, J. J., Bhakta, V., Brentnall, T. A., Luttges, J., Kloppel, G., and Lemoine, N. R. (2005). Expression of S100P and its novel binding partner S100PBPR in early pancreatic cancer. Am J Pathol *166*, 81-92.

Dvorak, H. F. (1986). Tumors: wounds that do not heal. Similarities between tumor stroma generation and wound healing. N Engl J Med *315*, 1650-1659.

El-Rifai, W., Moskaluk, C. A., Abdrabbo, M. K., Harper, J., Yoshida, C., Riggins, G. J., Frierson, H. F., Jr., and Powell, S. M. (2002). Gastric cancers overexpress S100A calcium-binding proteins. Cancer Res *62*, 6823-6826.

Farrell, C. M., and Lukens, L. N. (1995). Naturally occurring antisense transcripts are present in chick embryo chondrocytes simultaneously with the down-regulation of the alpha 1 (I) collagen gene. J Biol Chem *270*, 3400-3408.

Fiers, W., Beyaert, R., Declercq, W., and Vandenabeele, P. (1999). More than one way to die: apoptosis, necrosis and reactive oxygen damage. Oncogene *18*, 7719-7730.

Finco, T. S., Westwick, J. K., Norris, J. L., Beg, A. A., Der, C. J., and Baldwin, A. S., Jr. (1997). Oncogenic Ha-Ras-induced signaling activates NF-kappaB transcriptional activity, which is required for cellular transformation. J Biol Chem *272*, 24113-24116.

Foecking, M. K., and Hofstetter, H. (1986). Powerful and versatile enhancer-promoter unit for mammalian expression vectors. Gene *45*, 101-105.

Frith, M. C., Pheasant, M., and Mattick, J. S. (2005). The amazing complexity of the human transcriptome. Eur J Hum Genet *13*, 894-897.

Garber, K. (2004). Aspirin for cancer chemoprevention: still a headache? J Natl Cancer Inst *96*, 252-253.

Gottlob, K., Majewski, N., Kennedy, S., Kandel, E., Robey, R. B., and Hay, N. (2001). Inhibition of early apoptotic events by Akt/PKB is dependent on the first committed step of glycolysis and mitochondrial hexokinase. Genes Dev *15*, 1406-1418.

Grell, M., Douni, E., Wajant, H., Lohden, M., Clauss, M., Maxeiner, B., Georgopoulos, S., Lesslauer, W., Kollias, G., Pfizenmaier, K., and Scheurich, P. (1995). The transmembrane form of tumor necrosis factor is the prime activating ligand of the 80 kDa tumor necrosis factor receptor. Cell *83*, 793-802.

Greten, F. R., Eckmann, L., Greten, T. F., Park, J. M., Li, Z. W., Egan, L. J., Kagnoff, M. F., and Karin, M. (2004). IKKbeta links inflammation and tumorigenesis in a mouse model of colitis-associated cancer. Cell *118*, 285-296.

Griffiths-Jones, S. (2004). The microRNA Registry. Nucleic Acids Res *32*, D109-111.

Grossmann, J. (2002). Molecular mechanisms of "detachment-induced apoptosis--Anoikis". Apoptosis *7*, 247-260.

Guarguaglini, G., Battistoni, A., Pittoggi, C., Di Matteo, G., Di Fiore, B., and Lavia, P. (1997). Expression of the murine RanBP1 and Htf9-c genes is regulated from a shared bidirectional promoter during cell cycle progression. Biochem J *325 (Pt 1)*, 277-286.

Guo, F., Gopaul, D. N., and van Duyne, G. D. (1997). Structure of Cre recombinase complexed with DNA in a site-specific recombination synapse. Nature *389*, 40-46.

Gupta, R. A., and Dubois, R. N. (2001). Colorectal cancer prevention and treatment by inhibition of cyclooxygenase-2. Nat Rev Cancer *1*, 11-21.

Hamdane, M., David-Cordonnier, M. H., and D'Halluin, J. C. (1997). Activation of p65 NF-kappaB protein by p210BCR-ABL in a myeloid cell line (P210BCR-ABL activates p65 NF-kappaB). Oncogene *15*, 2267-2275.

Hamilton, D. L., and Abremski, K. (1984). Site-specific recombination by the bacteriophage P1 lox-Cre system. Cre-mediated synapsis of two lox sites. J Mol Biol *178*, 481-486.

Hanahan, D. (1983). Studies on transformation of Escherichia coli with plasmids. J Mol Biol *166*, 557-580.

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

Hansen, J. J., Bross, P., Westergaard, M., Nielsen, M. N., Eiberg, H., Borglum, A. D., Mogensen, J., Kristiansen, K., Bolund, L., and Gregersen, N. (2003). Genomic structure of the human mitochondrial chaperonin genes: HSP60 and HSP10 are localised head to head on chromosome 2 separated by a bidirectional promoter. Hum Genet *112*, 71-77.

Harris, T. J., and Peifer, M. (2005). Decisions, decisions: beta-catenin chooses between adhesion and transcription. Trends Cell Biol *15*, 234-237.

Hengartner, M. O. (2000). The biochemistry of apoptosis. Nature *407*, 770-776.

Hertel, S. C., Chwieralski, C. E., Hinz, M., Rio, M. C., Tomasetto, C., and Hoffmann, W. (2004). Profiling trefoil factor family (TFF) expression in the mouse: identification of an antisense TFF1-related transcript in the kidney and liver. Peptides *25*, 755-762.

Hsu, H., Shu, H. B., Pan, M. G., and Goeddel, D. V. (1996). TRADD-TRAF2 and TRADD-FADD interactions define two distinct TNF receptor 1 signal transduction pathways. Cell *84*, 299-308.

Huang, F. Y., Chiu, P. M., Tam, K. F., Kwok, Y. K., Lau, E. T., Tang, M. H., Ng, T. Y., Liu, V. W., Cheung, A. N., and Ngan, H. Y. (2006). Semi-quantitative fluorescent PCR analysis identifies PRKAA1 on chromosome 5 as a potential candidate cancer gene of cervical cancer. Gynecol Oncol *2*, 2.

Huang, T. T., Kudo, N., Yoshida, M., and Miyamoto, S. (2000). A nuclear export signal in the N-terminal regulatory domain of IkappaBalpha controls cytoplasmic localization of inactive NF-kappaB/IkappaBalpha complexes. Proc Natl Acad Sci U S A *97*, 1014-1019.

Hughes-Davies, L., Huntsman, D., Ruas, M., Fuks, F., Bye, J., Chin, S. F., Milner, J., Brown, L. A., Hsu, F., Gilks, B., *et al.* (2003). EMSY links the BRCA2 pathway to sporadic breast and ovarian cancer. Cell *115*, 523-535.

Hui, E. K., Wang, P. C., and Lo, S. J. (1998). Strategies for cloning unknown cellular flanking DNA sequences from foreign integrants. Cell Mol Life Sci *54*, 1403-1411.

Ji, J., Zhao, L., Wang, X., Zhou, C., Ding, F., Su, L., Zhang, C., Mao, X., Wu, M., and Liu, Z. (2004). Differential expression of S100 gene family in human esophageal squamous cell carcinoma. J Cancer Res Clin Oncol *130*, 480-486.

Kamata, H., Honda, S., Maeda, S., Chang, L., Hirata, H., and Karin, M. (2005). Reactive oxygen species promote TNFalpha-induced death and sustained JNK activation by inhibiting MAP kinase phosphatases. Cell *120*, 649-661.

Kapranov, P., Drenkow, J., Cheng, J., Long, J., Helt, G., Dike, S., and Gingeras, T. R. (2005). Examples of the complex architecture of the human transcriptome revealed by RACE and high-density tiling arrays. Genome Res *15*, 987-997.

Karin, M. (2005). Inflammation and cancer: the long reach of Ras. Nat Med *11*, 20-21.

Karin, M., and Ben-Neriah, Y. (2000). Phosphorylation meets ubiquitination: the control of NF-[kappa]B activity. Annu Rev Immunol *18*, 621-663.

Karin, M., and Greten, F. R. (2005). NF-kappaB: linking inflammation and immunity to cancer development and progression. Nat Rev Immunol *5*, 749-759.

Kastan, M. B., and Bartek, J. (2004). Cell-cycle checkpoints and cancer. Nature *432*, 316-323.

Katayama, S., Tomaru, Y., Kasukawa, T., Waki, K., Nakanishi, M., Nakamura, M., Nishida, H., Yap, C. C., Suzuki, M., Kawai, J., *et al.* (2005). Antisense transcription in the mammalian transcriptome. Science *309*, 1564-1566.

Kim, T. H., Barrera, L. O., Zheng, M., Qu, C., Singer, M. A., Richmond, T. A., Wu, Y., Green, R. D., and Ren, B. (2005). A high-resolution map of active promoters in the human genome. Nature *436*, 876-880.

Kiyosawa, H., Mise, N., Iwase, S., Hayashizaki, Y., and Abe, K. (2005). Disclosing hidden transcripts: mouse natural sense-antisense transcripts tend to be poly(A) negative and nuclear localized. Genome Res *15*, 463-474.

Kleinjan, D. A., and van Heyningen, V. (2003). Turned off by RNA. Nat Genet *34*, 125-126.

Knudsen, S. (1999). Promoter2.0: for the recognition of PolII promoter sequences. Bioinformatics *15*, 356-361.

Koller, E., Propp, S., Zhang, H., Zhao, C., Xiao, X., Chang, M., Hirsch, S. A., Shepard, P. J., Koo, S., Murphy, C., *et al.* (2006). Use of a chemically modified antisense oligonucleotide library to identify and validate Eg5 (kinesin-like 1) as a target for antineoplastic drug development. Cancer Res *66*, 2059-2066.

Kramps, T., Peter, O., Brunner, E., Nellen, D., Froesch, B., Chatterjee, S., Murone, M., Zullig, S., and Basler, K. (2002). Wnt/wingless signaling requires BCL9/legless-mediated recruitment of pygopus to the nuclear beta-catenin-TCF complex. Cell *109*, 47-60.

Krutzfeldt, J., Rajewsky, N., Braich, R., Rajeev, K. G., Tuschl, T., Manoharan, M., and Stoffel, M. (2005). Silencing of microRNAs in vivo with 'antagomirs'. Nature *438*, 685-689.

Kucharczak, J., Simmons, M. J., Fan, Y., and Gelinas, C. (2003). To be, or not to be: NF-kappaB is the answer--role of Rel/NF-kappaB in the regulation of apoptosis. Oncogene *22*, 8961-8982.

Kuwabara, T., Hsieh, J., Nakashima, K., Taira, K., and Gage, F. H. (2004). A small modulatory dsRNA specifies the fate of adult neural stem cells. Cell *116*, 779-793.

Kyriakis, J. M. (2001). Life-or-death decisions. Nature *414*, 265-266.

Lacroix, L., Lazar, V., Michiels, S., Ripoche, H., Dessen, P., Talbot, M., Caillou, B., Levillain, J. P., Schlumberger, M., and Bidart, J. M. (2005). Follicular thyroid tumors with the PAX8-PPARgamma1 rearrangement display characteristic genetic alterations. Am J Pathol *167*, 223-231.

Landry, J. R., Mager, D. L., and Wilhelm, B. T. (2003). Complex controls: the role of alternative promoters in mammalian genomes. Trends Genet *19*, 640-648.

Langley, K. E., Villarejo, M. R., Fowler, A. V., Zamenhof, P. J., and Zabin, I. (1975). Molecular basis of beta-galactosidase alpha-complementation. Proc Natl Acad Sci U S A *72*, 1254-1257.

Lemaire, P., Revelant, O., Bravo, R., and Charnay, P. (1988). Two mouse genes encoding potential transcription factors with identical DNA-binding domains are activated by growth factors in cultured cells. Proc Natl Acad Sci U S A *85*, 4691-4695.

Lewandowski, S. A., Thiery, J., Jalil, A., Leclercq, G., Szczylik, C., and Chouaib, S. (2005). Opposite effects of estrogen receptors alpha and beta on MCF-7 sensitivity to the cytotoxic action of TNF and p53 activity. Oncogene *24*, 4789-4798.

Li, Y. Y., Qin, L., Guo, Z. M., Liu, L., Xu, H., Hao, P., Su, J., Shi, Y., He, W. Z., and Li, Y. X. (2006). In silico discovery of human natural antisense transcripts. BMC Bioinformatics *7*, 18.

Liu, Y., Black, J., Kisiel, N., and Kulesz-Martin, M. F. (2000). SPAF, a new AAA-protein specific to early spermatogenesis and malignant conversion. Oncogene *19*, 1579-1588.

Lu, Q., Paredes, M., Medina, M., Zhou, J., Cavallo, R., Peifer, M., Orecchio, L., and Kosik, K. S. (1999). delta-catenin, an adhesive junction-associated protein which promotes cell scattering. J Cell Biol *144*, 519-532.

Maecker, H., Varfolomeev, E., Kischkel, F., Lawrence, D., LeBlanc, H., Lee, W., Hurst, S., Danilenko, D., Li, J., Filvaroff, E.*, et al.* (2005). TWEAK attenuates the transition from innate to adaptive immunity. Cell *123*, 931-944.

Maeda, S., Kamata, H., Luo, J. L., Leffert, H., and Karin, M. (2005). IKKbeta couples hepatocyte death to cytokine-driven compensatory proliferation that promotes chemical hepatocarcinogenesis. Cell *121*, 977-990.

Majewski, N., Nogueira, V., Bhaskar, P., Coy, P. E., Skeen, J. E., Gottlob, K., Chandel, N. S., Thompson, C. B., Robey, R. B., and Hay, N. (2004). Hexokinase-mitochondria interaction mediated by Akt is required to inhibit apoptosis in the presence or absence of Bax and Bak. Mol Cell *16*, 819-830.

Malumbres, M., and Barbacid, M. (2001). To cycle or not to cycle: a critical decision in cancer. Nat Rev Cancer *1*, 222-231.

Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

Markowitz, D., Goff, S., and Bank, A. (1988). A safe packaging line for gene transfer: separating viral genes on two different plasmids. J Virol *62*, 1120-1124.

Massague, J. (2004). G1 cell-cycle control and cancer. Nature *432*, 298-306.

Maul, G. G., Negorev, D., Bell, P., and Ishov, A. M. (2000). Review: properties and assembly mechanisms of ND10, PML bodies, or PODs. J Struct Biol *129*, 278-287.

Mayo, M. W., and Baldwin, A. S. (2000). The transcription factor NF-kappaB: control of oncogenesis and cancer therapy resistance. Biochim Biophys Acta *1470*, M55-62.

Medico, E., Gambarotta, G., Gentile, A., Comoglio, P. M., and Soriano, P. (2001). A gene trap vector system for identifying transcriptionally responsive genes. Nat Biotechnol *19*, 579-582.

Meltzer, P. S. (2005). Cancer genomics: small RNAs with big impacts. Nature *435*, 745-746.

Mialon, A., Sankinen, M., Soderstrom, H., Junttila, T. T., Holmstrom, T., Koivusalo, R., Papageorgiou, A. C., Johnson, R. S., Hietanen, S., Elenius, K., and Westermarck, J. (2005). DNA topoisomerase I is a cofactor for c-Jun in the regulation of epidermal growth factor receptor expression and cancer cell proliferation. Mol Cell Biol *25*, 5040-5051.

Micheau, O., Lens, S., Gaide, O., Alevizopoulos, K., and Tschopp, J. (2001). NF-kappaB signals induce the expression of c-FLIP. Mol Cell Biol *21*, 5299-5305.

Micheau, O., and Tschopp, J. (2003). Induction of TNF receptor I-mediated apoptosis via two sequential signaling complexes. Cell *114*, 181-190.

Milcarek, C., Price, R., and Penman, S. (1974). The metabolism of a poly(A) minus mRNA fraction in HeLa cells. Cell *3*, 1-10.

Miller, A. R., Skotzko, M. J., Rhoades, K., Belldegrun, A. S., Tso, C. L., Kaboo, R., McBride, W. H., Jacobs, E., Kohn, D. B., Moen, R., and et al. (1992). Simultaneous use of two retroviral vectors in human gene marking trials: feasibility and potential applications. Hum Gene Ther *3*, 619-624.

Miyamoto, S. (2004). RelA life and death decisions. Mol Cell *13*, 763-764.

Nagy, A. (2000). Cre recombinase: the universal reagent for genome tailoring. Genesis *26*, 99-109.

Nelson, D. E., See, V., Nelson, G., and White, M. R. (2004). Oscillations in transcription factor dynamics: a new way to control gene expression. Biochem Soc Trans *32*, 1090-1092.

Nickoloff, B. J., Ben-Neriah, Y., and Pikarsky, E. (2005). Inflammation and cancer: is the link as simple as we think? J Invest Dermatol *124*, x-xiv.

Nickoloff, B. J., Lingen, M. W., Chang, B. D., Shen, M., Swift, M., Curry, J., Bacon, P., Bodner, B., and Roninson, I. B. (2004). Tumor suppressor maspin is up-regulated during keratinocyte senescence, exerting a paracrine antiangiogenic activity. Cancer Res *64*, 2956-2961.

Noguchi, M., Miyamoto, S., Silverman, T. A., and Safer, B. (1994). Characterization of an antisense Inr element in the eIF-2 alpha gene. J Biol Chem *269*, 29161-29167.

Ohler, U., Harbeck, S., Niemann, H., Noth, E., and Reese, M. G. (1999). Interpolated markov chains for eukaryotic promoter recognition. Bioinformatics *15*, 362-369.

Okazaki, N., Kikuno, R., Ohara, R., Inamoto, S., Koseki, H., Hiraoka, S., Saga, Y., Kitamura, H., Nakagawa, T., Nagase, T.*, et al.* (2004). Prediction of the coding sequences of mouse homologues of FLJ genes: the complete nucleotide sequences of 110 mouse FLJ-homologous cDnas identified by screening of terminal sequences of cDNA clones randomly sampled from size-fractionated libraries. DNA Res *11*, 127-135.

Osipovich, A. B., White-Grindley, E. K., Hicks, G. G., Roshon, M. J., Shaffer, C., Moore, J. H., and Ruley, H. E. (2004). Activation of cryptic 3' splice sites within introns of cellular genes following gene entrapment. Nucleic Acids Res *32*, 2912-2924.

Ota, T., Suzuki, Y., Nishikawa, T., Otsuki, T., Sugiyama, T., Irie, R., Wakamatsu, A., Hayashi, K., Sato, H., Nagai, K.*, et al.* (2004). Complete sequencing and characterization of 21,243 full-length human cDNAs. Nat Genet *36*, 40-45.

Passegue, E., Jochum, W., Schorpp-Kistner, M., Mohle-Steinlein, U., and Wagner, E. F. (2001). Chronic myeloid leukemia with increased granulocyte progenitors in mice lacking junB expression in the myeloid lineage. Cell *104*, 21-32.

Passegue, E., Wagner, E. F., and Weissman, I. L. (2004). JunB deficiency leads to a myeloproliferative disorder arising from hematopoietic stem cells. Cell *119*, 431-443.

Pasteris, N. G., Nagata, K., Hall, A., and Gorski, J. L. (2000). Isolation, characterization, and mapping of the mouse Fgd3 gene, a new Faciogenital Dysplasia (FGD1; Aarskog Syndrome) gene homologue. Gene *242*, 237-247.

Perkins, N. D. (2004). Regulation of NF-kappaB by atypical activators and tumour suppressors. Biochem Soc Trans *32*, 936-939.

Pikarsky, E., Porat, R. M., Stein, I., Abramovitch, R., Amit, S., Kasem, S., Gutkovich-Pyest, E., Urieli-Shoval, S., Galun, E., and Ben-Neriah, Y. (2004). NF-kappaB functions as a tumour promoter in inflammation-associated cancer. Nature *431*, 461-466.

Prest, S. J., May, F. E., and Westley, B. R. (2002). The estrogen-regulated protein, TFF1, stimulates migration of human breast cancer cells. Faseb J *16*, 592-594.

Prest, S. J., Rees, R. C., Murdoch, C., Marshall, J. F., Cooper, P. A., Bibby, M., Li, G., and Ali, S. A. (1999). Chemokines induce the cellular migration of MCF-7 human breast carcinoma cells: subpopulations of tumour cells display positive and negative chemotaxis and differential in vivo growth potentials. Clin Exp Metastasis *17*, 389-396.

Prestridge, D. S. (1995). Predicting Pol II promoter sequences using transcription factor binding sites. J Mol Biol *249*, 923-932.

Reddig, P. J., and Juliano, R. L. (2005). Clinging to life: cell to matrix adhesion and cell survival. Cancer Metastasis Rev *24*, 425-439.

Reese, M. G. (2001). Application of a time-delay neural network to promoter annotation in the Drosophila melanogaster genome. Comput Chem *26*, 51-56.

Robey, R. B., and Hay, N. (2005). Mitochondrial hexokinases: guardians of the mitochondria. Cell Cycle *4*, 654-658.

Rogic, S., Ouellette, B. F., and Mackworth, A. K. (2002). Improving gene recognition accuracy by combining predictions from two gene-finding programs. Bioinformatics *18*, 1034-1045.

Russ, A. P., Friedel, C., Ballas, K., Kalina, U., Zahn, D., Strebhardt, K., and von Melchner, H. (1996). Identification of genes induced by factor deprivation in hematopoietic cells undergoing apoptosis using gene-trap mutagenesis and site-specific recombination. Proc Natl Acad Sci U S A *93*, 15279-15284.

Saccani, S., Pantano, S., and Natoli, G. (2001). Two waves of nuclear factor kappaB recruitment to target promoters. J Exp Med *193*, 1351-1359.

Sakon, S., Xue, X., Takekawa, M., Sasazuki, T., Okazaki, T., Kojima, Y., Piao, J. H., Yagita, H., Okumura, K., Doi, T., and Nakano, H. (2003). NF-kappaB inhibits TNF-induced accumulation of ROS that mediate prolonged MAPK activation and necrotic cell death. Embo J *22*, 3898-3909.

Salditt-Georgieff, M., Harpold, M. M., Wilson, M. C., and Darnell, J. E., Jr. (1981). Large heterogeneous nuclear ribonucleic acid has three times as many 5' caps as polyadenylic acid segments, and most caps do not enter polyribosomes. Mol Cell Biol *1*, 179-187.

Sanchez-Elsner, T., Gou, D., Kremmer, E., and Sauer, F. (2006). Noncoding RNAs of trithorax response elements recruit Drosophila Ash1 to Ultrabithorax. Science *311*, 1118-1123.

Schernhammer, E. S., Kang, J. H., Chan, A. T., Michaud, D. S., Skinner, H. G., Giovannucci, E., Colditz, G. A., and Fuchs, C. S. (2004). A prospective study of aspirin use and the risk of pancreatic cancer in women. J Natl Cancer Inst *96*, 22-28.

Schmidt, C., Fischer, G., Kadner, H., Genersch, E., Kuhn, K., and Poschl, E. (1993). Differential effects of DNA-binding proteins on bidirectional transcription from the common promoter region of human collagen type IV genes COL4A1 and COL4A2. Biochim Biophys Acta *1174*, 1-10.

Schuh, R., Aicher, W., Gaul, U., Cote, S., Preiss, A., Maier, D., Seifert, E., Nauber, U., Schroder, C., Kemler, R., and et al. (1986). A conserved family of nuclear proteins containing structural elements of the finger protein encoded by Kruppel, a Drosophila segmentation gene. Cell *47*, 1025-1032.

Schwamborn, J., Lindecke, A., Elvers, M., Horejschi, V., Kerick, M., Rafigh, M., Pfeiffer, J., Prullage, M., Kaltschmidt, B., and Kaltschmidt, C. (2003). Microarray analysis of tumor necrosis factor alpha induced gene expression in U373 human glioblastoma cells. BMC Genomics *4*, 46.

Schwartz, F., Maeda, N., Smithies, O., Hickey, R., Edelmann, W., Skoultchi, A., and Kucherlapati, R. (1991). A dominant positive and negative selectable gene for use in mammalian cells. Proc Natl Acad Sci U S A *88*, 10416-10420.

Screaton, R. A., Kiessling, S., Sansom, O. J., Millar, C. B., Maddison, K., Bird, A., Clarke, A. R., and Frisch, S. M. (2003). Fas-associated death domain protein interacts with methyl-CpG binding domain protein 4: a potential link between genome surveillance and apoptosis. Proc Natl Acad Sci U S A *100*, 5211-5216.

Shendure, J., and Church, G. M. (2002). Computational discovery of sense-antisense transcription in the human and mouse genomes. Genome Biol *3*, RESEARCH0044.

Silverman, T. A., Noguchi, M., and Safer, B. (1992). Role of sequences within the first intron in the regulation of expression of eukaryotic initiation factor 2 alpha. J Biol Chem *267*, 9738-9742.

Sitcheran, R., Gupta, P., Fisher, P. B., and Baldwin, A. S. (2005). Positive and negative regulation of EAAT2 by NF-kappaB: a role for N-myc in TNFalpha-controlled repression. Embo J *24*, 510-520.

Soule, H. D., Vazguez, J., Long, A., Albert, S., and Brennan, M. (1973). A human cell line from a pleural effusion derived from a breast carcinoma. J Natl Cancer Inst *51*, 1409-1416.

Sovak, M. A., Bellas, R. E., Kim, D. W., Zanieski, G. J., Rogers, A. E., Traish, A. M., and Sonenshein, G. E. (1997). Aberrant nuclear factor-kappaB/Rel expression and the pathogenesis of breast cancer. J Clin Invest *100*, 2952-2960.

Sparmann, A., and Bar-Sagi, D. (2004). Ras-induced interleukin-8 expression plays a critical role in tumor growth and angiogenesis. Cancer Cell *6*, 447-458.

Stanford, W. L., Cohn, J. B., and Cordes, S. P. (2001). Gene-trap mutagenesis: past, present and beyond. Nat Rev Genet *2*, 756-768.

Steel, D. M., and Harris, H. (1989). The effect of antisense RNA to fibronectin on the malignancy of hybrids between melanoma cells and normal fibroblasts. J Cell Sci *93 (Pt 3)*, 515-524.

Strausberg, R. L., Feingold, E. A., Grouse, L. H., Derge, J. G., Klausner, R. D., Collins, F. S., Wagner, L., Shenmen, C. M., Schuler, G. D., Altschul, S. F., *et al.* (2002). Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences. Proc Natl Acad Sci U S A *99*, 16899-16903.

Strausberg, R. L., Feingold, E. A., Klausner, R. D., and Collins, F. S. (1999). The mammalian gene collection. Science *286*, 455-457.

Stumpo, D. J., Byrd, N. A., Phillips, R. S., Ghosh, S., Maronpot, R. R., Castranio, T., Meyers, E. N., Mishina, Y., and Blackshear, P. J. (2004). Chorioallantoic fusion defects and embryonic lethality resulting from disruption of Zfp36L1, a gene encoding a CCCH tandem zinc finger protein of the Tristetraprolin family. Mol Cell Biol *24*, 6445-6455.

Sun, G., Liu, X., Mercado, P., Jenkinson, S. R., Kypriotou, M., Feigenbaum, L., Galera, P., and Bosselut, R. (2005). The zinc finger protein cKrox directs CD4 lineage differentiation during intrathymic T cell positive selection. Nat Immunol *6*, 373-381.

Sutterluety, H., Bartl, S., Doetzlhofer, A., Khier, H., Wintersberger, E., and Seiser, C. (1998). Growth-regulated antisense transcription of the mouse thymidine kinase gene. Nucleic Acids Res *26*, 4989-4995.

Tada, Y., Wada, M., Migita, T., Nagayama, J., Hinoshita, E., Mochida, Y., Maehara, Y., Tsuneyoshi, M., Kuwano, M., and Naito, S. (2002). Increased expression of multidrug resistance-associated proteins in bladder cancer during clinical course and drug resistance to doxorubicin. Int J Cancer *98*, 630-635.

Tang, G., Minemoto, Y., Dibling, B., Purcell, N. H., Li, Z., Karin, M., and Lin, A. (2001). Inhibition of JNK activation through NF-kappaB target genes. Nature *414*, 313-317.

Teratani, T., Watanabe, T., Kuwahara, F., Kumagai, H., Kobayashi, S., Aoki, U., Ishikawa, A., Arai, K., and Nozawa, R. (2002). Induced transcriptional expression of calcium-binding protein S100A1 and S100A10 genes in human renal cell carcinoma. Cancer Lett *175*, 71-77.

Thellin, O., Zorzi, W., Lakaye, B., De Borman, B., Coumans, B., Hennen, G., Grisar, T., Igout, A., and Heinen, E. (1999). Housekeeping genes as internal standards: use and limits. J Biotechnol *75*, 291-295.

Thiefes, A., Wolter, S., Mushinski, J. F., Hoffmann, E., Dittrich-Breiholz, O., Graue, N., Dorrie, A., Schneider, H., Wirth, D., Luckow, B., *et al.* (2005). Simultaneous blockade of NFkappaB, JNK, and p38 MAPK by a kinase-inactive mutant of the protein kinase TAK1 sensitizes cells to apoptosis and affects a distinct spectrum of tumor necrosis factor [corrected] target genes. J Biol Chem *280*, 27728-27741.

Thorey, I. S., Muth, K., Russ, A. P., Otte, J., Reffelmann, A., and von Melchner, H. (1998). Selective disruption of genes transiently induced in differentiating mouse embryonic stem cells by using gene trap mutagenesis and site-specific recombination. Mol Cell Biol *18*, 3081-3088.

Tian, B., Nowak, D. E., and Brasier, A. R. (2005). A TNF-induced gene expression program under oscillatory NF-kappaB control. BMC Genomics *6*, 137.

Tian, B., Nowak, D. E., Jamaluddin, M., Wang, S., and Brasier, A. R. (2005). Identification of direct genomic targets downstream of the nuclear factor-kappaB transcription factor mediating tumor necrosis factor signaling. J Biol Chem *280*, 17435-17448.

Ting, S. B., Wilanowski, T., Cerruti, L., Zhao, L. L., Cunningham, J. M., and Jane, S. M. (2003). The identification and characterization of human Sister-of-Mammalian Grainyhead (SOM) expands the grainyhead-like family of developmental transcription factors. Biochem J *370*, 953-962.

Tremain, R., Marko, M., Kinnimulki, V., Ueno, H., Bottinger, E., and Glick, A. (2000). Defects in TGF-beta signaling overcome senescence of mouse keratinocytes expressing v-Ha-ras. Oncogene *19*, 1698-1709.

Trinklein, N. D., Aldred, S. F., Hartman, S. J., Schroeder, D. I., Otillar, R. P., and Myers, R. M. (2004). An abundance of bidirectional promoters in the human genome. Genome Res *14*, 62-66.

Tufarelli, C., Stanley, J. A., Garrick, D., Sharpe, J. A., Ayyub, H., Wood, W. G., and Higgs, D. R. (2003). Transcription of antisense RNA leading to gene silencing and methylation as a novel cause of human genetic disease. Nat Genet *34*, 157-165.

van 't Veer, L. J., Dai, H., van de Vijver, M. J., He, Y. D., Hart, A. A., Mao, M., Peterse, H. L., van der Kooy, K., Marton, M. J., Witteveen, A. T., *et al.* (2002). Gene expression profiling predicts clinical outcome of breast cancer. Nature *415*, 530-536.

von Melchner, H., Reddy, S., and Ruley, H. E. (1990). Isolation of cellular promoters by using a retrovirus promoter trap. Proc Natl Acad Sci U S A *87*, 3733-3737.

von Melchner, H., and Ruley, H. E. (1989). Identification of cellular promoters by using a retrovirus promoter trap. J Virol *63*, 3227-3233.

Voziyanov, Y., Pathania, S., and Jayaram, M. (1999). A general model for site-specific recombination by the integrase family recombinases. Nucleic Acids Res *27*, 930-941.

Wajant, H., Pfizenmaier, K., and Scheurich, P. (2003). Tumor necrosis factor signaling. Cell Death Differ *10*, 45-65.

Wan, Y., and Nordeen, S. K. (2002). Identification of genes differentially regulated by glucocorticoids and progestins using a Cre/loxP-mediated retroviral promoter-trapping strategy. J Mol Endocrinol *28*, 177-192.

Wang, C. Y., Mayo, M. W., Korneluk, R. G., Goeddel, D. V., and Baldwin, A. S., Jr. (1998). NF-kappaB antiapoptosis: induction of TRAF1 and TRAF2 and c-IAP1 and c-IAP2 to suppress caspase-8 activation. Science *281*, 1680-1683.

Wang, L. H. (2004). Molecular signaling regulating anchorage-independent growth of cancer cells. Mt Sinai J Med *71*, 361-367.

Weber, F., de Villiers, J., and Schaffner, W. (1984). An SV40 "enhancer trap" incorporates exogenous enhancers or generates enhancers from its own sequences. Cell *36*, 983-992.

Weinberg, R. A. (1996). How cancer arises. Sci Am *275*, 62-70.

Weiss, P. A. (1971). A cell is not an island entire of itself. Perspect Biol Med *14*, 182-205.

Wempe, F., Yang, J. Y., Hammann, J., and von Melchner, H. (2001). Gene trapping identifies transiently induced survival genes during programmed cell death. Genome Biol *2*, RESEARCH0023.

Widom, R. L., Culic, I., Lee, J. Y., and Korn, J. H. (1997). Cloning and characterization of hcKrox, a transcriptional regulator of extracellular matrix gene expression. Gene *198*, 407-420.

Wilkinson, D. G., Bhatt, S., Chavrier, P., Bravo, R., and Charnay, P. (1989). Segment-specific expression of a zinc-finger gene in the developing nervous system of the mouse. Nature *337*, 461-464.

Willingham, A. T., Orth, A. P., Batalov, S., Peters, E. C., Wen, B. G., Aza-Blanc, P., Hogenesch, J. B., and Schultz, P. G. (2005). A strategy for probing the function of noncoding RNAs finds a repressor of NFAT. Science *309*, 1570-1573.

Wilson, C., Bellen, H. J., and Gehring, W. J. (1990). Position effects on eukaryotic gene expression. Annu Rev Cell Biol *6*, 679-714.

Wu, X., Li, Y., Crise, B., and Burgess, S. M. (2003). Transcription start regions in the human genome are favored targets for MLV integration. Science *300*, 1749-1751.

Yang, M. Y., Liu, T. C., Chang, J. G., Lin, P. M., and Lin, S. F. (2003). JunB gene expression is inactivated by methylation in chronic myeloid leukemia. Blood *101*, 3205-3211.

Yelin, R., Dahary, D., Sorek, R., Levanon, E. Y., Goldstein, O., Shoshan, A., Diber, A., Biton, S., Tamir, Y., Khosravi, R., *et al.* (2003). Widespread occurrence of antisense transcription in the human genome. Nat Biotechnol *21*, 379-386.

Zamore, P. D., and Haley, B. (2005). Ribo-gnome: the big world of small RNAs. Science *309*, 1519-1524.

Zamoyska, R. (2005). Switching sides: transcriptional control of CD4 thymocyte lineage commitment. Nat Immunol *6*, 363-364.

Zhou, A., Scoggin, S., Gaynor, R. B., and Williams, N. S. (2003). Identification of NF-kappa B-regulated genes induced by TNFalpha utilizing expression profiling and RNA interference. Oncogene *22*, 2054-2064.

7. - Summary

One of the hallmarks of cancer is the escape of the transformed cells from apoptosis. Therefore, the identification of survival genes, allowing cancer cells to circumvent programmed cell death, could provide new diagnostic markers as well as targets for therapeutic intervention. A well known transcription factor regulating the balance between pro- and anti- apoptotic factors is NF-κB, which is strongly induced by tumor necrosis factor alpha (TNFα). When cells are stimulated by TNFα their response is biphasic with an initial NF-κB induction of survival genes which is overridden by the subsequent activation of initiator caspases triggering apoptosis.

By combining gene trap mutagenesis with site specific recombination a strategy was developed, which enriches for genes induced by TNFα in the human breast cancer cell line MCF-7. The strategy relies on a one way gene expression switch based on Cre/*loxP* mediated recombination, which uncouples the expression of a marker gene from the trapped cellular promoter thereby enabling the recovery of genes that are only transiently induced by TNFα. The marker gene used in these experiments was a dominant negative variant of the TNFα-receptor associated protein FADD (dnFADD), which blocks the apoptotic branch of the TNFα induced signaling pathway.

Initial experiments indicated that MCF-7 cells expressing high levels of dnFADD were insensitive to TNFα induced apoptosis and therefore suitable for the installment of a one way gene expression switch susceptible to Cre/*loxP* mediated recombination. A MCF-7 reporter clone harboring the recombinase dependent gene expression switch was infected with the gene trap retrovirus U3Cre, which inserts the *Cre* recombinase gene into a large collection of chromosomal sites. Insertion of *Cre* downstream of an active cellular promoter induces dnFADD expression from the gene expression switch enabling the cells to block TNFα triggered apoptosis.

From a gene trap integration library containing approximately $2x10^6$ unique proviral integrations, 69 unique TNFα inducible gene trap insertion sites were recovered in a two step selection procedure. Sequencing of the genomic regions adjacent to the insertion sites, which were obtained by inverse PCR (gene trap sequence tags, GTSTs), and data base analysis

revealed that 42% of the GTSTs belonged to annotated genes, 13% to known cDNAs with open reading frames, 17% to Genscan predicted genes, 9% to ESTs, 9% to repetitive sequences and 10% to unannotated genomic sequence. Overall, 44% of the annotated genes recovered in this screen were directly or indirectly related to cancer, indicating that the gene trap strategy developed here is suitable for the identification of cancer relevant genes.

Analysis of the expression patterns of the trapped and annotated genes in wild type cells revealed that 19 out of 24 genes were either up- or down- regulated by a factor of at least 1.45 by TNFα.

A large fraction of the gene trap insertions were located upstream, in introns or in opposite orientation to annotated transcripts, indicating that the strategy efficiently recovers non-coding RNAs (ncRNAs). While the biological significance of these transcripts still needs to be elucidated, they fall into two main categories. The first category includes gene trap insertions upstream of genes, which could either represent regulatory RNAs interacting with promoter elements or transcripts driven by bidirectional promoters. The second includes inverse orientation gene trap insertions in introns of annotated genes suggesting the presence of natural antisense transcripts (NATs). Interestingly, more than 50% of all antisense integrations are located downstream of transcription start sites predicted by different algorithms supporting the existence of RNAs transcribed from the corresponding genomic regions. Intronic integrations on the coding strand could be derived from cryptic splicing, alternative promoter usage or additional, so far uncharacterized transcripts.

Preliminary functional analysis of two genes recovered in this screen encoding the transcription factor ZFP67 and the FLJ14451 protein revealed that *FLJ14451* but not *ZFP67* inhibited anchorage independent growth in soft agar, suggesting that *FLJ14451* might have some tumor suppressor functions.

In summary, besides identifying a putative tumor suppressor protein, the present experiments have shown that gene trapping is useful in identifying non-coding transcripts in living cells and may turn out to be the method of choice in characterizing these transcripts whose functions are still largely unknown.

8. - Zusammenfassung

Krebszellen sind nicht unabhängig von ihrer Umgebung, sondern stehen vielmehr im Fokus vielfältiger Interaktionen mit anderen Zelltypen, darunter Stroma-, Gefäß- und Immunzellen. Die Reaktion des Körpers gegenüber Tumoren weist viele Parallelen zur Wundheilung und zu entzündlichen Prozessen auf. Die Bedeutung dieser inflammatorischen Reaktionen spiegelt sich wider in dem Vergleich der, der Krebsentstehung zugrunde liegenden, genetischen Schädigung mit einem "Funken, der ein Feuer entzündet" und der Gleichsetzung mancher Entzündungsarten mit "Öl, das ins Feuer gegossen wird". Entzündliche Prozesse wurden mit allen Stufen der Tumorentstehung und -progression in Verbindung gebracht, dabei stellte sich der Transkriptionsfaktor NF-κB als Bindeglied zwischen Krebs und Entzündungen heraus. In malignen Zellen reguliert NF-κB nach seiner Aktivierung durch Onkogene oder inflammatorische Zytokine antiapoptotische und proliferative Gene. Einer der stärksten bekannten Induktoren NF-κB abhängiger Genexpression in malignen und inflammatorischen Zellen ist Tumornekrosefaktor alpha (TNFα), ein Zytokin, das in vielen Tumoren produziert wird. Aufgrund der Besonderheiten des TNFα-Signalweges kann TNFα sowohl pro-, als auch antiapototische Effekte haben. Zellen, die TNFα Rezeptor 1 (TNFR1) exprimieren, zeigen eine biphasische Reaktion auf TNFα. Nach einer initialen, NF-κB vermittelten Induktion von Überlebensgenen kommt es zur Modifikation des TNFR-Signalkomplexes und Rekrutierung des Adaptor-Proteins FADD ("Fas associated death domain protein"). Dies führt zur Aktivierung von Initiator-Caspasen und letztendlich zur Einleitung des programmierten Zelltods. Der Beweis dafür, dass die primär aktivierten Gene antiapoptotische Funktionen haben, wurde durch die simultane Behandlung von Zellen mit TNFα und Translationsinhibitoren erbracht, welche den apoptotischen Prozess auslösen oder beschleunigen können. Die Deregulation solcher Überlebensgene ist für die Apoptoseresistenz von Tumorzellen, ein zentrales Charakteristikum von Krebserkrankungen, verantwortlich. Daher ist die Identifizierung dieser Gene von besonderen Interesse für der Krebsforschung, da ihre Genprodukte als diagnostische Marker oder als Angriffspunkte für neuartige Therapieansätze dienen könnten.

In dem hier beschriebenen Projekt wurde die Herausforderung, TNFα regulierte Überlebensgene zu identifizieren, mit einer Genfallenstrategie angegangen. Genfallen sind auf

Plasmiden oder Retroviren basierende Vektoren, mit Hilfe derer ein Reportergen zufällig über das Genom verteilt inseriert werden kann. Die hier verwendete Genfalle U3Cre enthält ein *Cre*-Rekombinasegen anstelle eines konventionellen Reporters. Die Kombination dieser Genfalle mit einem Cre-abhängigen, irreversiblen, molekularen Schalter ermöglicht die Identifizierung von transient exprimierten Genen.

Dieser Schalter besteht aus einem konstitutiv aktiven Promoter 5'-oberhalb von zwei Reportergenen, die beide ein eigenes Polyadenylierungs-(polyA-)Signal tragen. Die 5'-Kassette kodiert für eine Neomycin-Phosphotransferase (neo), die Neomycin-Resistenz vermittelt. Sie ist von zwei gleichartig orientierten *loxP*-Sequenzen flankiert und verhindert aufgrund ihrer Polyadenylierung die Transkription der 3' dazu liegenden Kassette. Diese enthält ein Gen für eine dominant negative Variante des FADD Proteins (dnFADD), welche in der Lage ist, TNFR-vermittelte Apoptose zu blockieren. Im Grundzustand exprimiert das Schalterkonstrukt neo, aber nicht dnFADD. Cre-induzierte Rekombination führt zur Deletion der neo-Kassette und damit verbunden zur Transkription von *dnFADD*, da dieses nun unter die Kontrolle des Promotors kommt. Dadurch verlieren die Zellen ihre Neomycin-Resistenz und auch, aufgrund der Expression von *dnFADD*, die Sensitivität gegenüber TNFα-induzierter Apoptose.

Die erste zu klärende Frage war, ob die für das Vorhaben ausgewählte Brustkrebszellinie MCF-7 nach *dnFADD*-Expression resistent gegenüber TNFα wird. Vorexperimente zeigten, dass extrem hohe dnFADD-Level zur TNFα-Resistenz führen. Daher wurden mit einer retroviralen Transduktion MCF-7 Reporterzellklone generiert, welche den molekularen Schalter zur konditionalen Expression von *dnFADD* tragen. Diese Klone wurden auf dnFADD-Protein-Level und TNFα-Resistenz nach Cre-induzierter Rekombination untersucht. Ein Klon, der die gewünschten Charakteristika aufwies wurde zur Erzeugung der U3Cre Genfallen-Integrationsbank benutzt. Um eine repräsentative Verteilung der Genfallenintegrationen über das Genom zu erreichen, wurden die Bedingungen so eingestellt, dass eine Integrationsbank mit ungefähr $2x10^6$ unabhängigen proviralen Integrationen erzeugt wurde. Um aus dieser Bank Zellen mit Genfallen-Integrationen in TNFα-induzierbaren Genen anzureichen, wurde eine zweistufige Selektionsstrategie eingesetzt. Im ersten Schritt wurden in Abwesenheit von TNFα Zellen mit Integrationen in konstitutiv aktiven Genen eliminiert. Die zweite Selektion in Gegenwart des Zytokins diente der Anreicherung von

Zellen mit Insertionen in TNFα-induzierten Genen. Aus diesen Selektionsprozeduren wurden 78 Zellklone erhalten und in diesen die mit der Genfalle markierten Gene identifiert. Dazu wurden die flankierend zu den proviralen Sequenzen liegenden genomischen Abschnitte (GTSTs, "gene trap sequence tags") in einer inversen PCR amplifiziert und sequenziert. Von den so erhaltenen 69 unterschiedlichen GTSTs lagen 42% in annotierten Genen, 13% in Genen mit offenen Leserahmen unbekannter Funktion, 17% in hypothetischen Genen, die mit dem Genscan-Algorithmus vorhergesagt wurden, 9% in ESTs ("expressed sequence tags"), 9% in repetitiven Elementen und 10% in nicht annotierten genomischen Regionen. 44% der aus diesem Screening-Verfahren erhaltenen, bekannten Gene liessen sich direkt oder indirekt mit Krebserkrankungen korrelieren. Dies ist ein Indiz dafür, dass der hier entwickelte, experimentelle Ansatz zur Identifizierung Krebs-relevanter Gene geeignet ist.

Die U3Cre Genfalle basiert auf einem Maus Moloney Leukämie Virus (MMLV), der präferentiell in die 5'-Enden von Genen inseriert. Ein zusätzlicher Faktor, der diese Präferenz verstärken könnte, ist ein Stopcodon im gleichen Leseraster wie das AUG des *Cre*-Gens, was zu einem starken Selektionsdruck für Integrationen führt, in denen das *Cre* AUG das erste Initiationscodon im enstehenden Fusionstranskript ist. Da die Genfalle zudem keinen Spleißakzeptor enthält, ist sie konzeptionell eine Exon-Genfalle. All dies führte zu der Erwartung, dass die Mehrzahl der Integrationen in 5'-Exons der getroffenen Gene liegen sollte.

Überraschenderweise wurden auch andersgeartete Integrationen erhalten. Generell konnten die Genfallen-Insertionen in verschiedene Gruppen unterteilt werden, welche sich bezüglich der Position der Genfalle innerhalb des getroffenen Gens und ihrer Orientierung relativ zur Transkriptionsrichtung des Gens unterschieden.

Integrationen in oder zwischen Exons auf dem kodierenden Strang ("sense") von Genen weisen auf die Entstehung eines Fusionstrankripts aus dem 5'-Ende einer mRNA und dem Genfallen-kodierten Cre hin. Die häufig beobachteten Integrationen in Introns, die auch für vergleichbare Exon-Genfallen beschrieben worden waren, könnten das Resultat der Aktivierung kryptischer Spleißakzeptoren 5' zum Genfallen-Provirus sein. Eine alternative Erklärung wäre die Insertion der Genfalle in unabhängige Transkriptionseinheiten innerhalb eines Introns. "Sense"-Insertionen oberhalb von Genen können als Indiz für die Aktivierung benachbarter, alternativer Promotoren angesehen werden können, welche im Genom

weitverbreitet sind, was in einer kürzlich veröffentlichten Studie belegt wurden. In dieser konnten im statistischen Mittel jedem 3'-Ende von Genen 1.32 Transkriptionsstarts zugeordnet werden.

Ein verblüffender Befund bei der Analyse der Genfallen-Integrationen war der hohe Prozentsatz an "antisense"-Integration, von ca. 40%. Im Gegensatz zu den bisher geschilderten Kategorien deuten diese Insertionen auf die Existenz von Promotoren auf dem nicht-kodierenden Strang hin, welche die Transkription von "antisense"-Transkripten kontrollieren, die möglicherweise regulatorische Funktionen haben. Eine Analyse mit mehreren frei verfügbaren Promotor-Vorhersageprogrammen ergab, dass mehr als 50% dieser Integrationen mit potentiellen Transkriptionsstarts 5'-oberhalb der proviralen Insertionsstellen assoziiert sind. "Antisense"-Integrationen könnten die Aktivität bidirektionaler Promotoren widerspiegeln, die die Transkription zweier gegenläufig angeordneter Gene steuern, oder ein Hinweis auf die Existenz natürlicher "antisense"-Transkripte (NATs) sein. Beispiele individueller, bidirektionaler Genpaare sind seit langem bekannt. Diese Anordnung kann dazu dienen stöchiometrische Verhältnisse zwischen Proteinen aufrecht zu erhalten oder Gene, deren Produkte in einem gemeinsamen Signalweg benötigt werden gemeinsam zu regulieren. Zudem können so auch zeitlich regulierte Transkriptionsprogramme kontrolliert und zelluläre Antworten auf äußere Stimuli, wie z.B. TNFα, koordiniert werden. Natürliche "antisense"-Transkripte, auf der anderen Seite, sind mit vielen Aspekten eukaryotischer Genregulation in Verbindung gebracht worden, darunter Imprinting, RNA-Interferenz und –Editing, sowie der Inaktivierung eines der beiden X-Chromosomen bei weiblichen Säugetieren. Es wäre daher nicht überraschend, wenn "antisense"-RNAs auch Teil des von TNFα aktivierten regulatorischen Netzwerks wären.

Die Frage, um welchen Typ von Transkripten es sich bei den RNAs handelt, die von bisher nicht charakterisierten Loci ausgehen und mit der Genfalle identifiziert wurden, kann nicht schlüssig beantwortet werden. Eine vor kurzem veröffentlichte Transkriptom-Analyse deutete darauf hin, dass polyadenylierte Transkripte wesentlicher häufiger vorkommen als erwartet und dass diese RNAs den Großteil der gesamten RNA im Zellkern und Zytoplamsa ausmachen. Normalerweise werden nicht polyadenalierte RNAs sehr schnell abgebaut, sie könnten aber durch die von dem Genfallenvektor ausgelöste Polyadenylierung stabilisiert werden.

Bedingt durch die hohe Sensitivität der hier eingesetzten Genfallen-Strategie ist es wahrscheinlich, dass neue Transkripte identifiziert wurden, die sich mit anderen Nachweistechniken nicht detektieren lassen. Dies wird unterstützt durch drei GTSTs, die in nicht annotierten Regionen des menschlichen Genoms liegen, in den aber zumindest von dem Genscan-Algorithmus Gene vorhergesagt werden. Da Gen-Voraussageprogramme wie Genscan auch strukturelle Eigenschaften mit einbeziehen, wie z.B. das Vorhandensein von offenen Leserahmen mit Initiations- und Terminationscodons, entgehen ihnen nicht-kodierende RNAs ("non-coding" RNAs, ncRNAs). Genfallen-Integrationen in genomische Regionen ohne jegliche Annotation könnten somit durchaus in Gene für ncRNAs erfolgt sein. Die bisher beschriebenen ncRNAs haben alle regulatorische Funktionen. Unter diesen Funktionen ist auch die epigenetische Translationssuppression durch micro RNAs (miRNAs), welche in den letzten Jahren Gegenstand intensiver Untersuchungen geworden sind. Eine Genfallen-Integration befand sich 10 kbp 3' zu einer bekannten miRNA (mo-mir21), was nahe legt, dass in diesem Fall eine prä-miRNA, die oft sehr große, im Kern lokalisierte Vorstufe der micro RNAs, "abgefangen" wurde.

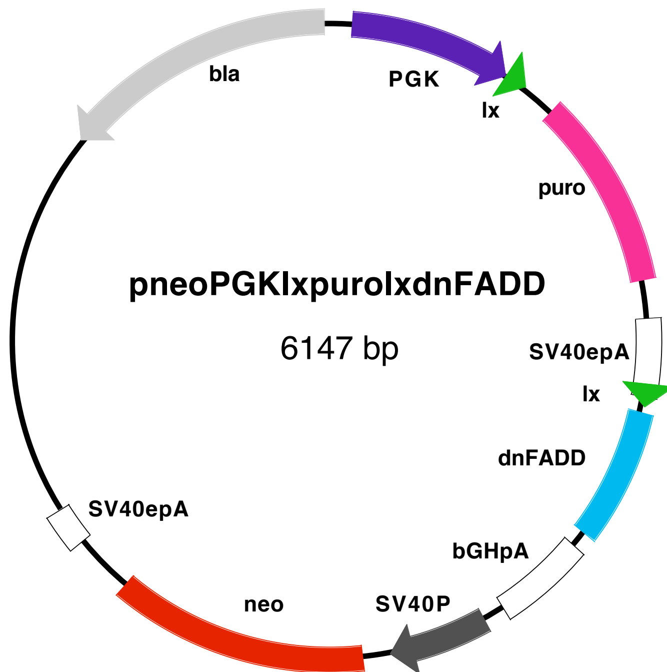
Um Gene für funktionelle Ananlysen auszuwählen, wurde zunächst die Regulation von 24 GTSTs durch TNFα in Northern Blots untersucht. Dabei erwiesen sich 5 Gene als nicht reguliert, während die 1 anderen 9 eine Hoch- oder Herunterregulation um einen Faktor von mindestens 1.45 zeigten. Kandidatengene für funktionelle Untersuchungen wurden anhand der Regulation durch TNFα, einer möglichen Assoziation mit Krebserkrankungen und der Verfügbarkeit von cDNAs mit kompletten offenen Leserastern ausgewählt. Eine vorläufige funtkionelle Analyse von zwei dieser Kandidaten, dem Transkriptionsfaktor ZFP67 und dem FLJ14451 Protein, ergaben, dass *FLJ14451*, aber nicht *ZFP67*, in der Lage ist, das Substrat-unabhängige Wachstum von MCF-7 Zellen in Weichagar zu inhibieren. Dies weist auf eine mögliche Tumorsuppresorfunktion des Proteins hin.

Zusammenfassend lässt sich sagen, dass die im Rahmen dieser Dissertation durchgeführten Experimente nicht nur zur Identifizierung eines potentiellen, neuen Tumorsuppressorgens führten, sondern auch zeigten, dass Genfallen ein nützliches Werkzeug bei der Suche nach nicht-kodierenden RNAs in lebenden Zellen sein können und ihr Einsatz möglicherweise die Methode der Wahl für die Identifizierung derartiger Transkripte darstellt.

9. - Annex

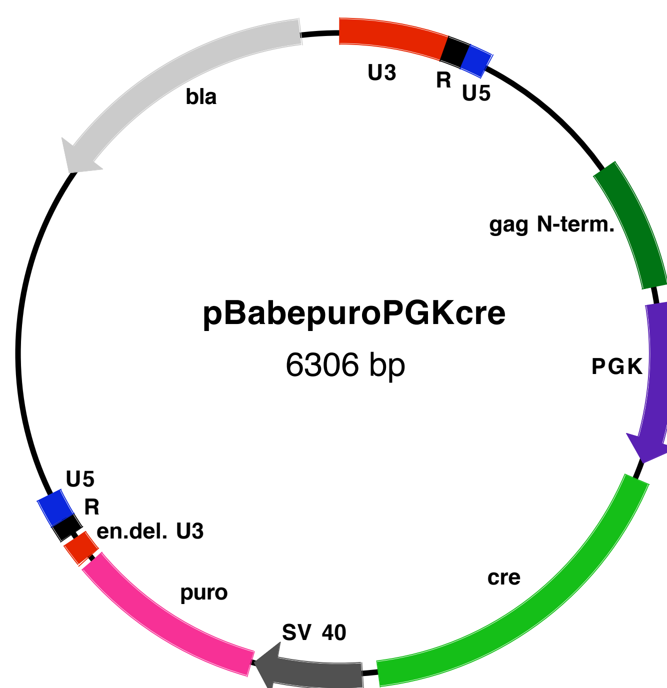
9.1. Plasmids maps

Plasmid containing the PGKpurodnFADD switch vector



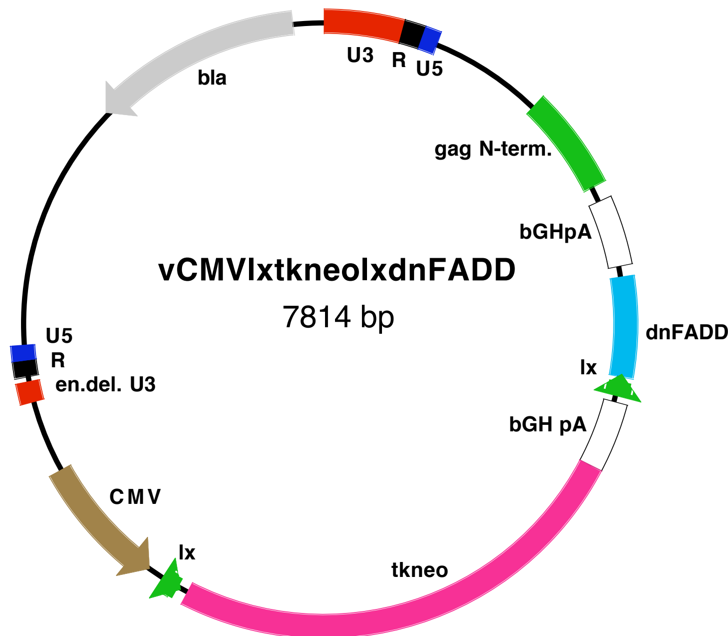
bla: beta lactamase conferring ampicillin resistance
neo: neomycin-phosphotransferase conferring G418 resistance
puro: puromycin-acetyl-transferase conferring puromycin resistance
PGK: promoter from the mouse *phosphoglycerate-kinase* gene
SV40P: promoter from the Simian virus 40 (SV40)
bGHpA: bovine growth hormone polyadenylation sequence
SV40epa: SV40 early polyadenylation sequence
dnFADD: human 5' truncated dominant negative FADD
lx: *loxP* sites from bacteriophage P1

Plasmid for forced recombination of the switch vector



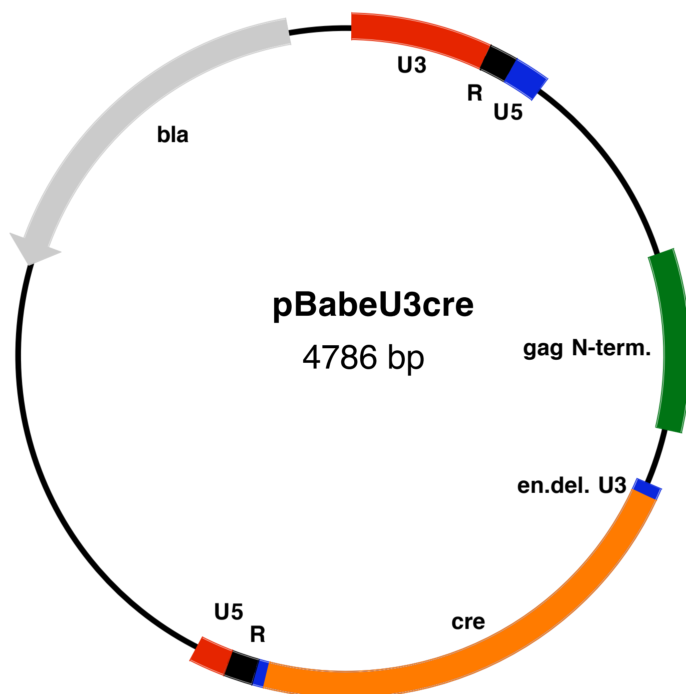
bla: beta lactamase conferring ampicillin resistance
puro: puromycin-acetyl-transferase conferring puromycin resistance
PGK: promoter from the mouse *phosphoglycerate-kinase* gene
bGHpA: bovine growth hormone polyadenylation sequence
cre: *Cre* recombinase gene from bacteriophage P1
U3: U3 region from the Moloney murine leukemia virus
U5: U3 region from the MMLV
En.del.U3: enhancer and promoter deleted U3 region from the MMLV
R: R region from the MMLV l
gag N-term: truncated open reading frame gag necessary as extended packaging signal

Plasmid containing the CMVtk-neodnFADD switch vector



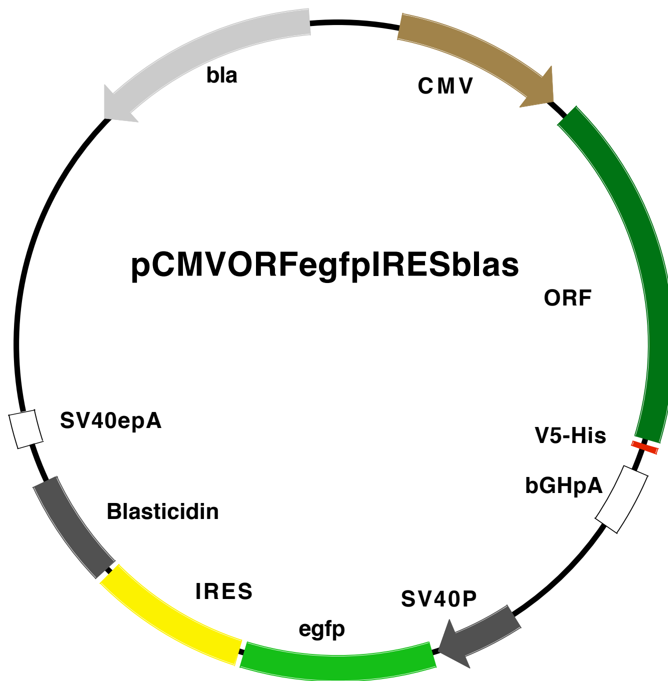
bla: beta lactamase conferring ampicillin resistance
tkneo: fusion between timidin-kinase and neomycin-phosphotransferase conferring G418 resistance and gancyclovir sensitivity
CMV: promoter from cytomegalovirus immediate early region
bGHpA: bovine growth hormone polyadenylation sequence
dnFADD: human 5' truncated dominant negative FADD
lx: *loxP* sites from bacteriophage P1
U3: U3 region from the MMLV
U5: U3 region from the MMLV
En.del.U3: enhancer and promoter deleted U3 region from the MMLV
R: R region from the MMLV l
gag N-term: truncated open reading frame gag necessary as extended packaging signal

Plasmid used for the gene trap virus production



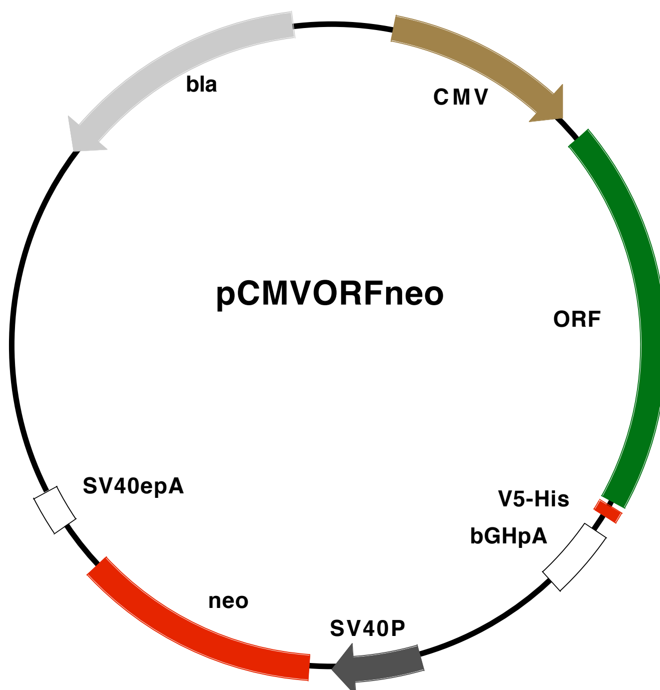
bla: beta lactamase conferring ampicillin resistance
gag N-term: truncated open reading frame gag necessary as extended packaging signal
U3: U3 region from Moloney murine leukemia virus (MMLV)
U5: U5 region from the MMLV
En.del.U3: enhancer and promoter deleted U3 region from the MMLV
R: R region from the MMLV long terminal repeat region (LTR)
cre: *Cre* recombinase gene from bacteriophage P1

Plasmid used for ORF over-expression



bla: beta lactamase conferring ampicillin resistance
Blasticidin: blasticidin deaminase conferring blasticidin resistance
CMV: promoter from the cytomegalovirus immediate early region
SV40P: promoter from the Simian virus 40 (SV40)
bGHpA: bovine growth hormone polyadenylation sequence
SV40epa: SV40 early polyadenylation sequence
EGFP: enhanced green fluorescent protein
ORF: Open reading frame
IRES: internal ribosomal entry site
V5: V5-his epitope

Plasmid used for ORF over-expression



bla: beta lactamase conferring ampicillin resistance
neo: neomycin-phosphotransferase conferring G418 resistance
CMV: promoter from the cytomegalovirus immediate early region
SV40P: promoter from the Simian virus 40 (SV40)
bGHpA: bovine growth hormone polyadenylation sequence
SV40epa: SV40 early polyadenylation sequence
ORF: Open reading frame
V5-his: V5-his epitope

9.2. Primers

9.2.1. For probe amplification

These primers were used in RT-PCRs to amplified short regions to obtain gene specific probes for Northern blot.

Gene	Primer name		Sequence		sonde length
ABCC3	MOAT D for	5'-	GTCGCCCTGCCCTGCTACTTG	-3'	319 bp
	MOAT D rev	5'-	AGGTGGTGAAGCGGAAGGGGT	-3'	
BCL9L	BCL9 for 2	5'-	AGCCAATGCACCCAGAAAATAA	-3'	492 bp
	BCL9 rev 2	5'-	CTGGCACGCTGCTCTCGCTGAG	-3'	
C1QTNF6	TNF6 for1	5'-	TAGTATTGCAGACATGGGCCAAGG	-3'	820 bp
	TNF6 rev1	5'-	AAGTCGTTGCTGTAGATGGCGTTC	-3'	
c20orf142	c20orf142	5'-	GTTGTCCCCGTTGCCCGAGA	-3'	433 bp
	c20orf142	5'-	AGGGTGGTGATGGCGGTGTG	-3'	
CTNND2	CTNND2 for	5'-	GGGTCCGCGTAATTGGAGG	-3'	238 bp
	CTNND2 rev	5'-	GGGTCCGCGTAATTGGAGG	-3'	
EED	EED for 2	5'-	TGAGCAGTGACGAGAACAGCA	-3'	347 bp
	EED rev 2	5'-	TTCATCAGCATCAGCATCCAC	-3'	
ESR1	ESR-1 for	5'-	TGCCCTACTACCTGGAGAAC	-3'	650
	ESR-1 rev	5'-	GCCCATCATCGAAGCTTCAC	-3'	
FGD3	FGD3 for 2	5'-	TGCTGGGCAAGGAAGAGATTTT	-3'	586 bp
	FGD3 rev 2	5'-	GGGGTGTTCTCAGAGTCAGGTT	-3'	
FLJ14451	NM_032786 for	5'-	GCAGAACCCAGCAGTGATGT	-3'	787 bp
	NM_032786 rev	5'-	CAAAATGAGGGCCATCAGGG	-3'	
HexII	Hex for	5'-	CTGCTTGCCTACTTCTTCACG	-3'	351 bp
	Hex rev	5'-	CACTGCCTCGCATGATGTCCT	-3'	
JunB	Jun B for	5'-	AGCCCTTCTACCACGACGACT	-3'	533 bp
	Jun B rev	5'-	GGTTGGTGTAAACGGGAGGTG	-3'	

Gene	Primer name		Sequence		sonde length
KTCD5	FLJ20040 for	5'-	GAGTGTTGGAGGAAGCAGAAT	-3'	220bp
	FLJ20040 rev	5'-	GTACAAAGTTCTCTTGCCTGG	-3'	
RNF184	QLHC17 for	5'-	AATGAACCCCGTGAATGCTAC	-3'	331 bp
	QLHC17 rev	5'-	TTATATCCCGTGCCAGTGTAG	-3'	
S100A10	S100A10 for	5'-	CTTCAACGGACCACACCAAA	-3'	346 bp
	S100A10 rev	5'-	TTATCAGGGAGGAGCGAACT	-3'	
S100PBP	FLJ12903 for	5'-	GCCCCAGCTCTCTTCTTCAAA	-3'	621 bp
	FLJ12903 rev	5'-	TGCTGATGGGATGACACAAAT	-3'	
SOM	Q8TE85 for	5'-	GTGCGGCTGCTAAAGAACGA	-3'	331 bp
	Q8TE85 rev	5'-	CAAATCTGGGTACTCTGGGG	-3'	
TFF1	TFF1 for	5'-	AACAAGGTGATCTGCGCCCT	-3'	195 bp
	TFF1 rev	5'-	GGGACGTCGATGGTATTAGG	-3'	
Top1	TOP I for 2	5'-	ACACAAAGATCGAGAACACCG	-3'	339 bp
	TOP I rev 2	5'-	AGGAGGAACAAAATAGCCATC	-3'	
WDR10	WDR10 for 2	5'-	GACACCTCTGATGGCACCTTA	-3'	458 bp
	WDR10 rev 2	5'-	CGCCATTTTTGTTCCGTATGC	-3'	
ZFP67	ZFP67 for	5'-	ACGGCTGAGAGGAGAAGATG	-3'	661 bp
	ZFP67 rev	5'-	CAGGGACTAGGTGGTTTGCT	-3'	
ZNF143	ZNF 143 for	5'-	CGTGGCAGATGGTGACAACTT	-3'	368 bp
	ZNF 143 rev	5'-	AGATGGTGTCAGACTGCGGGA	-3'	
ZNHIT2	C11 orf 5 for	5'-	CATACTCTCGCCCTGTATCACG	-3'	245 bp
	C11 orf 5 rev	5'-	GTAGCCTTTCTGGTTGGTCGG	-3'	
ZNRF1	NIN283	5'-	AAACCTCGCCTCTCCTACAAC	-3'	320 bp
	NIN283	5'-	GGAAGGGAGAACCATGATCTG	-3'	
β-actin	actin for	5'-	TCGAGCACGGCATCGTCACCAACT	-3'	551 bp
	actin rev	5'-	ACCGCTCATTGCCAATGGTGATGA	-3'	

9.2.2. For ORF amplification

These primers were used for amplification of ORF required for ORF expression vectors.

Gene		Forward			Reverse	
FLJ14451	5'-	AGCAGGATCCACCatgcct gaccgggacagctatgcc	-3'	5'-	GGTAGCTCTAGAgtgtgg catggccgtgatgcgcat	-3'
ZFP67	5'-	AGCAGGATCCACCatgggg agccccgaggatgacctg	-3'	5'-	GGTAGCTCTAGAagagga ctccatggcaccttcagc	-3'

9.2.3. For recombination test

These primers were used to analyze successful recombination in the reporter switch vector.

Primer		Sequence	
CMV for 2	5'-	GTACGGTGGGAGGTCTATATAAGCAG	-3'
FADD rev 4	5'-	GATGCTGTCGATCTTGGTGTCTGA	-3'

9.2.4. For inverse PCR

These primers were used for inverse PCR.

Primer		Sequence	
Cre 43	5'-	CGGTCAGTAAATTGGACACCTTCC	-3'
SY2	5'-	GCTAACTAGCTCTGTATCTGGCGGAC	-3'
CreiPCRrev2	5'-	GAGTGAACGAACCTGGTCGAAATCAG	-3'
iPCRU3	5'-	CCTCCGATTGACTGAGTCGCCC	-3'
Cre1P	5'-	GCATGCTAGCTTGCCAAACC	-3'
SY1	5'-	ACCCGTGGTGGAACTGACG	-3'
CreiPCRrev1	5'-	AACAGCATTGCTGTCACTTGGTCG	-3'
iPCRU4	5'-	TACCCGTGTATCCAATAAACCC	-3'

9.3. Sequences obtained by inverse PCR

The genomic DNA sequences flanking the U3Cre integration site are listed without the U3Cre-DNA junction and together with their clone identification (clone ID).

Clone ID	GTSTs without the U3Cre-DNA junction
MI3K01	GTAGAACCCACTGCATCACAACTGGAACTGATCTCTGGTTATGCCAATCTAGGTCTGCCTGGCAG GGTGAGTCTACCGCTCGAAAGCTGCACAGAGGAAATCCTCCCTCCTCCAGGTTCAGGCCCGGAGT TACGAGCGGAGATTCGCGGCGGCGGGTCAGAGCCCAGAGCAAACCAGGCTGGGGAACGCACCTCG CCCGCGGCTCCGGAAGCGGCTGCGGCCCCGACCGGAAAGCCCCCACATCGCCCGAAGGAATCGCG GCTGGCCTGCGG
MI3K02	CCTGCATTACTCAGCAGGCAACTGGCTTCCCTCACCCCACCCCAAACTGAGAAGCCTCAACTGTC AGCGGATCTGACTTCAAATCCAAACCATTAGGGAGGAGAAAAGCAAGAAACCTTTCTGTTCAAGT AACCCAGCTTCAGAAACACATACCCCCTAACGTGCAGTAATGAATAGATTCTTGTCTGTAGCCTC CGCAGTTTTAGCCAGCAAATTTAAGTGGGATGGAGACACATGTCAAGA
MI3K03	GACCAGAAAGAGATGGAGGAAGAGCAGGCTTGATGAAGGATCTTCCAATCATAGATTAATTTGGC CCGGTGCAGTGAAACAGACAGACGTGCCTTTGGTACTTGGTTCACCACCTGTGCTTTGGCCGAAG GAATGGCGCTGCCCCTAGCCGTGGACACTGCTGCCCTGTCCAGCCTCAGTCCAGCTTTGCCTAAA TGAAACCAGGAACTCTGC
MI3K04	GAGGATGCAGTCATTGGCTATGGTGGTTGTGGCTTCCTATGTGTGTGGAGGAGGGCTGCAGGGAG GAGGGGTGTCAGGGAA
MI3K05	ATCAGACCCTTGGTGACCCAGGGCTACATGTACTCGGCGGAAAACTGGGGTTAAACCGGACCCTT GGTTGGTTGAAAAATTAAAATAATTTTTAAAAAGTGGGCGGTGAAATAATCCAGAAGTTGGAGGG CCTGGACAAAGGCTGGGGAGGGG
MI3K09	GGGTGCACTTTTATCATAAACTCCAGCAGCCTTCAGTTTTCCATCGTTAAGTATAAATGTACGTG TTGTATATTTTTTCTGTGCTCCAGTGGACTCGTCGACATTATATAACTTAAAAGCATTTGGAGTG AGATGGAAAGTAGGAGTTGGCTGTTTCTCACTCAGAAATGTTGGATTCAGCTGATCCTCCTCCTG AAACTTTTGAACTTAAAGAGCAGGTTTGGGTCATTTCCTTTGCAAAAAGCAGCCTTAGG
MI3K10	CCCAACACCCCCCACTCACTCATGTGCAAGGGGCCCACTCTCGTTTTAGTC
MI3K11	GACCAGTAAGAGGGAGGAAGAGCAGGCTTGATGAAGGATCTTCCAATCATAGATTAATTTGGCCC GGTGCAGTGAAAAGACAGACGTGCCTTTGGTACTTGGTTCACCACCTGTGCTTTGCCGAAGAATG CGCTCCCCTAGCCGTGGACACTGCTGCCTGTCCAAGCCTCAGTCAGCTCTGCCTAAGTGAAACCA GGAGCTCTGCTTTGTGGCTTGGGGTGAATCAGAGGTTAAGCAAGCCAGCCAAGGGAGCCCTAAGC CAGATTTTGGAAGCCTTAAGGCCACCCCGTTTTTTGCCCCAACCAGTTCCGTTTTCGGGAGTTTT TG
MI3K12	GCCTGGTTCCAAGGAGAGCTCCCTGGGAAGCAAAGGCTGAGAGAAGCGCTCTCTCTCTCTCTTTC TGCATCTGCCGGGAGACTGCAGAGGACAAGAACGCGCGGGTTT
MI3K13	AGGGCAGGCAGCCTGACTTCATTTCTCGCCTGAACAAGGACCATGCTGTCCTGCACGCTGGGTCT GACCGTCTGCCCTCTCTCCCCAGCACCAAGCGTGACCTTGGCTGTGGCGCTCAACGGCCAGCTCC GGCGGCCCCTCTGCTGCTCCTCGGCTTTCCCGGAAGTGGGAGAAGCCTGCCTGGCCTCGGCCTTT GTCCAACGACA
MI3K14	TTTCCCCAGGGACACAAACACTGCGGAAGGCCGCAGGGTCCTCTGCCTAGGAAAACCAGAGACCT TTGTGCACTTGTTTATCTGCTGACCTTCCCTCCACTATTGTCCTATGACCCTGCCAAATCCCCCT CTGCGAGAAACACCCAAGAATGATCAATAAAAAAAAAAAAAAAAAAA

Clone ID	GTSTs without the U3Cre-DNA junction
MI3K16	CTCTGGGCGAGGCGACAGCGGGCCGGCTCGGGCGGGGACAGCGGAGGACCGGGGTGCACTCACAC GTTGAGGACGTTGCGCTTGTTGCTGAGGTAGCTCTCGGGCAACGGGGACAACTCCTTGAGGAGGG AGCCCGCCAGCATGGAGGCCACCAGGGCCCAGGGCAGGTAGCGCCGCACGGCCGCCCGCACCAGC GTCCCCCGCAACAACCACTCGCAGCGCTCCAGATGCTCCATGCCGGATCTCGTCAGCCACCGTCC TCCTCTCCGTGCCCTCTCGGCCACCGTATCGCCCTTCGCCCGGACCTGCGCCTCCACCTCCCTTC GAGCGCATGCGCACTGCCCGCGA
MI3K17A	GCGCTGACCCAGCTTTTCTAACGAATTCGACCCTGTGTACCCCTTGACCCGCTAACGGTCGGCAG GCTTTGACTGCT
MI3K18	GTATGTGCCCTCCAAAATTCCATGTTGGAACTTAATATCCTATGTAATGGTATTAAGAGGTGGGG CCTTTGGAGAAGTGAGTAGGTTCTGAACATCCCGCCCTTATGAACAGGATTAGGTGTTCTTATAG AAGAGGTTGATGGGAGTGCCTAGTCCCTTGTGCCCTTCTGTCTTCCCCTATGTGAGGCTGTAGAA ACAAGATGCCATATTAGTAGCAGAGAGTAAATCTTCACCAGACACTACATCTGCTGGCACCTTGA CCTTTTACTTT
MI3K21	CCGAGCACGTGGTCGGCTCCGCGGGGGTCTTCCGGCGTCCCGGGGTGGGGGACTTGTGGGTGCCT TTACCGCAGCCCATTGGACCAATCCGGCATGCACTTCCTTCCCCTTCTTGAGGCTTCATAAAAAG CCCCTTGGGACTTCAAGCCAAGGAGCAGGAGAGGAAGCCAAGTT
MI3K22	TTTTCACTGTGAGTCACTTCTGTTTCTATCCACCCTAGGGGATTTGTTCTCCATGCGATCAGGGT GGACAGCTCACGGCTTCCCTTGAGTGGATTTCATTTTAGGGGCCCTCACTCCCCCAACCTCTGCC TTCACGCCATCTTGAATTGCCACAGTGTGCTAGTCCAGAAACCCCACACAGCGCACGATCAGCCT TACCATCCAGGGTGAGGGAGGAAAAAGAAGCTGTTAAAAAGTCATCTTTCGGATTTAAATCAGCC CTGCTGAAATAATCGAGGGAACTCGGAACTATGGAAAGGtCAAAGGTTTGCAGGTTTAGGGTTCG TGG
MI3K23	AGACCACCATGTCCAGCTAATTTTTCTAGTTTTTAGTTTAGTCCGGCGTTTCTCCATGTTAGGCC AGGCTCGGTCTTGAACTCCTGCCTCAAGTGACTCGCCCACCTCAGCCTCCCAAAGTGCTGGGATT ACAGGCCTGAACCACTGTGCCTGCCGAGATATCCCATTTTGGATAGGAGCCTGGGACATATTCTC AACTTCCGTGGGTCTGCCACTACTTATCAGAAGTTTCTGCTCACCTTACACAACTGTGAGTGATG CCCAGGGAGCAGCCTCAGTGGCCCAAGGGGAAAAACCTCTGGATGAAGCCCTGACCCCTCACATC TCAGCCTCCATCACTGCCCCTGCTACCCAAAGTCAAGCCCAGAAACCAGATGTGAGCCAAACCCG GCAGCCCATAAAGG
MI3K24	GTCCCAGCTACTCGGGAGGATGAGAAGAAGAATTGCTTGAACCTGGAGGCAGAGGTTCACGGAGC CAAGATTGTAGCCACTCGCTGCCCAGCCTGTACAACAGAGCAAGACTCCATCTCAAAAAAAAAAA A
MI3K36	AAGCACTTATGCAAGTGCGGGGCACAGAGAAAATTCTGGTAACACTCCTCCCCTCAATTCTGATC AGGTCCCAGTCTCAACTAATCAGTTCAGGGATGGGCCCTTTGCCACCTTTTTCCTAACCAAAGCC ACTCTCAAGGGGAAACTTCTATTAACGGCCAGTAAGCAAGACACACAAC
MI3K37	GATTGACACCCGTCAGCGGGGTCTTTCAGGGAAGCAGGCAGCCCGGGGGAGCCCCGAAAAATCAG CAGAGAGCAGCCCTGGCTTCTCAGAGGAGCGAGGAGTAGCTGCTGAGGGGAGGAGGGGAGCAGAG TGGTCAAGTTTGCATAGTCATTGTCCCCATGCCTCCTTGCAAGGCAGGCCAGGGAGCCGTGGGCC TCTCGCGCCCTGCTTAGCCAGTCCTGCAGGCAAAATCAACGTGGGTTCCTGTCGCCCTGCCACGA TGCCAAGGGGGACCCCTCTCGGTCCTTACCCTAAGAAAAACCCAAAGCCCTCCTCGCAGGCC
MI3K39A	CAAGTTACTGATGTAGTTGTTACGACCAATCTTTCATACTTCTTGGTTAAGAATCTGTCCGGTTC TAAAGAGTGCATTTCATATCCTTGCTAAGCCTACTAATAAGCTTCATCCCTTTTTTTTTTTTTTT TTTT
MI3K40	CTAGTAGCTAAGTCCCCACCTCCTTGCCCTCCGTG

Clone ID	GTSTs without the U3Cre-DNA junction
MI3K44	GATTTAATCATATAATTATGAATCTGTTCTTTTTTTCCCCCCAAATATTTGTAGCTTTAGGTAGT AGTTACCAGAATGATGAATTTTCCTCCTATGCTCCGTAGTCTTGTAATAAAAAGCATGTACAGTG TAGACGTTTGCTAGGCATGGCTCTTCCTTTGACCAGTTCATGAACTGGAGTCAGGGTGGGGGTGA AGGCTAGGGTCCGAGGTCTGGGTCAGGCGCTCGTGCACCGTCGTGGGAGGTCGAGGCTGATGTCC GCTGCTCCCCATCAGCGTTAGTGCCCAGATCACCCCGTGCTCCGTCCCGCCGGGAAGCG
MI3K47	CTGTGAGCTCAGTCCAAAACTTGCTACAGGTAAGCCTGTGGGCAAGTCACATGGGTTGTCTCGGG CCTTGGTTTCCCCTTTCGAGGAGGGAGAAGGAGGGAGGTCCCCTGTGTAAAGTCTCCCAGCACCA GTTCAATGTTCTTTTCTCCTATTCATGCCACCCAAAGTAGGGGGATCATGTCCAGAAATGCTCGT GCCTTTCACCAGTGTTATAGGGGTCCTTTGCCAGGGGACCTTCAACCAGCTGAGGGCATGTGTCT AGGGACTGCCCTCAAGTGTGGAAAATACCTGGGCTTCCTTCCCCTTCCTAGG
MI3K61	GATCCTGGGCAGGTCGACACACATGCCAGCTGGGCCGACCCTGTACTTCCCTGGCCCTGGGGGGA CAAGGGAAATTGCAAGTGAATCGCTCGCCAGGTTATCAGGGGGTAACCATTTCCTGGAGAGCCCA AAGGCCTGGGACTCAAGAAGGGCCTG
MI3K62	CCAGAAGAGTGGAGGAAGAGAGCTTGTGAAGGATCTTCCAATCATAGATTAATTTGGCCCGGTGC AGTGAAAAAGACAGACGTGCCTTTGGTACTTGGTTCACCACCTGTGCTTTGCCGAAGAAATGGCG CTCCCTAGCCGTGGACACTTGCTTGCCCTGTCCAGCCTCAGTCCAGCTCTTGCCTTAAGTGAAAC AGGAGCTCTGCCTTGTTGCTTGGGGGTGGAATGCAGGAGCGCTCCAAGCAGGCCAAGCCAGGGAG CCCTAAGCCAGATTCTGGGATGCATTAAGCCACCCCGTTTCTGCCCCAACCAAGTCC
MI3K66	GGGATGAGCTGGTTCCAGGAGAAAGGCCGCGAGCCCGTTTGGGGGCCTGGCCGCGGTTTTTTCT
MI3K67	ACACTTCACACACAACCCCCAGCCCACCTGCTCCAGGCTTTGGGTGGGGGGCGGAGGGTACTGCC TGGGAGCTGCATTTGGGGGTTCACCTGAGGGTTGTCATCTGTTAGCACTGACTCTGGTTGTTTCT CCC
MI3K76	GTATGTATCTCAGTCTTGATTTAAGAAACACTTTTTTTTCAACTGGAGAATACAGTTTAATTTTG TGTTACCCCATGATACCATTTTATTAGTACGATTTTCTTCAAGGACTCAAAAATAGCTCTGTTCA TTAATCTGAGCACACAGGTAACTGCACATTAGTACAGTCCAACTAACTGAGCTTGTCTTGCATAT GGAAGACACATCCTTAACAACTAACGTCAAATACTGGCAATGTTCTAACTACATACAACACCGTT CAGCTGGAAGTGCAGTCACCAGCTTCTCACTTCTTACATCCAATCCCATCCTAGTTGTCCTTCTG CTCTCAAAA
MI3K80	CTGAGGAAGTACAGGAGGGGTGAATTCCACCATGGAGAAAGTATTCAATAACAAGCCTACAGAGA AGTTGAGGAAGGTGGAGGAGTAAAGGACAATGATGTCCTCGAGAAACCTGCTCATGAAACCTGCT CAAGAAACCTGAGGAGGCAAAAGCTGATTCTAAAGTGTTGAGAGACAACAGCACAAACCTGGG
MI3K81	GTAGGGCTGTTGGAAGGATGGAAGGAGTTCAGAAACAGAAAGCTCCCAGCACAGAGCCTGGCCGT AATAAACAGTCTATGTGTTTGCTATCACGAGGGTACTTTTAAAGAAATTTAGGTGCAGGATGAGT ATAAATGAGACGTGGAGTGAGTAGCAGAATCAACACAAGGTTTTTCTTATAATGGATGAGAGATT GTAAAACACATTTGCATTGGGGAGAAGGACA
MI7K07	AGGGCTGGAGGTGGGCATGCGGCAGATTACTGTTGTAAAGCATTGAGATGTTTATGTGTATGCAT ATCTAAAAGCACAGCACTTAATCCTTTACATTGTCTATGATGCAAAGACCTTTGTTCACGTGTTT GTCTTGCTGACCTCTCCCCACAA
MI7K08	AAAGGAAAAAAAAAAAAGGTTGGTCCTGGCCGGTGCTGTGGCTCACGCCTGTAATCCTAGCATTTG GGAGG
MI7K20	GGGAAAAAAAAGCCTGGTGTGGTGATGAGTCTTTTGGTCCCATCTACTCGGGATGCTGAGGAAGAA GGATCGCTGAACACGGGAGA
MI7K26	GGCTTCGGGGTGGAATCTCACGATCGAAAGGAAAGTGGGCCCAAGGTTGGGTGGAGGCTCCCTAG TAGTTTTAAGCCCCCCCGTCTCCAAAGCACTTCAATTTTCTTCTCCCCATCCAGGTTCAGGACCT GGGT

Clone ID	GTSTs without the U3Cre-DNA junction
MI7K27B	CCCTACACAGGGCTCCTAGCCATAAGCACTTATGCAAGTGCGGGGCACAGAGAAAATTCTGGTAA CACTCCTCCCCTCAATTCTGATCAGGTCCCAGGCTCAACTAATCAGTTCAGGGATGGGCCCTTTG CCACCTTTTTCCTAACCAAAGCCACTCTCAAGGGGAAACTTCTATTAACGGCC
MI7K28	CCGAGACCTGGGAACTGCTAGCATTCTGTTTCTCCTGCTTCTGGGTTATGCAGCAGGAAGAGAGG GCATGGAGTTTGGAGCCACACAGACCTTAGGTTCAAATCCAGCTCTGTGACTTTGAAAAGTTACT TAACATCTCTGAGCCTTAGTTTCATCATGCAAAGATAGGACATTTCTCCTGAGGTGGTTGTGAGG GTTAAATAAAGTAAGTCAAGAGCCCAAGACATAATAGGAAAGCAATGAATCGTCTCTATTATTAA TTTTTATTTTTATATATCAAAGTCTGGGGTACAGTGGCCTTGCTTAAAAATGGAATCCCCTTTCC TTTGAATTACTTTTTCTCCCCATATGATAGTCTCAAATATTTATTCATTTTTCTTTTCTTTCTTT TTTTTTTT
MI7K29	AAGGAGTATCCTTTCCACCACCGCCTCGAGACGAGGCGAAACTCAGGGGATGCTCCCTCACCCCC TTCCTCCTTCCAAAGACGTTTCAGCCGCAATTGGCCCGTTCCTTCGATCTCCGACCCGCTGCTAG GGAGCGAAAGTCTGTGGAATGGGCTACACCACCAATGTCCCAGGAAACGGGGGAAGATACAAGAC GTCACCTCTTGCTTT
MI7K30B	GTGAGCCTTCCTCTCCTGCTGAAATTCCGGCGGCTTGGCAACCGGCCGGGGGGTCTTGGATTCCT CGGGGAGACACCACTGATGCTTTGTGGTTTCACGTAATTTGGATTTAAAAGTTAA
MI7K31	GTTACAAAGCAGGACTTCCTCATAGGGAGTAAAGTGTTTTCTTTTCATAAATATAACCTATTTAT TTATATTTTCAAGTAATCTATTGCAGCCTTAGTTGTAGTAGGTCACTAGTTCATGCCTTCAATAA ACTGTCATTTAGTGCTTACTGTGTGTGAGGCATTGCACTAGTAAAGAAAGAGGATATAAAACAAA AATCACTGCATAGAAAATAAAGTGCAGTGGGAATTTGGAAACAAGCTGTTTTTT
MI7K32	AGACGGGCCTCCTAAAGCACTGGGATTATAGGCGTGAGCC
MI7K34	ATAAATGATTACATTTGTAGGACATTTTGGTGCCTAATGCCAGTCAGCAAGGGTTGGTTGCACAA TGAGTTTGGACAGG
MI7K48	GTGTTAGTAAATGCATACTATTTTGTCTGTATTTAAATCGAGTCCAAATCTCTCGCTCTACAGCC CGCCTTGGGATGTTTCTTATATCCCAAGAAACAGAATATTTTGATGGGATCGCTGATGTTTCAGA CTGCAAAAGCAGCTCAGGGCGTTTCAGTCGTGAAGTCAACAAGATAACCGTCTGGACCGGAAGCT GGCTCCTCCCGGTCT
MI7K49	GACCTCCTGGCCACATGGTGAAAGCCCGTCTCTACTAAAAATACAAAAATTAGCTGGGCGTGGTG GCACGTGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAGGAGAAGTGCTTGAACCCAGGAGGC AGAGGTTGCAGTGAGCCCAAGATCTTGCCACTGCACTCCAGCCAGGGCAGAAGAACGAGACTCCG TCTCAAAAAAAAAAAAAAAAAAAAAAAGGGTTTTTTTAAAAAATT
MI7K50	GTAACGTGACTATGAGATTCTACAACTGTAAAGCACCCTTGCTCCTTCAATGATGATTTACCGAG CACGGATTTTGCTATAGCCCAGTACCAGTGTGCCTACCCCGGGGCTCGCAGCCAGGTTAGGAGAG CAACCAGCACTCCAGCAGGGACCACATAGCAAGACCCCATCTGTACTAAGAATAAAACAAATTA
MI7K51	CGGACATGTTCTTAGAATCCCCCACTACTATCTGATGATGTTAAGCATCTTTATTTATTTATTTT GAGAGCTCTGCTCTGCACCCC
MI7K53	GATGCCTCCGCCGCGGAGCCCGCGGCCGCCGAGCGGGTTCTTGGAGATGTCCCGGGGGCCTGCAC GCCCGTCGTACCCACCCGCATCCCCGCGATAGTCAGCCTGAGCCGCGGCCCAGTCTCGCCGCTCG TGCGCTTCCAGCTGCCCAATGTGCTGTTCGCCTACGCGCATACTCTCGCCCTGTATCACGGCGGT GACGACGCGCTGCTCTCTGACTTCTGTGCCACACTGCTCGGCGTTTCCGGAGCCCTGGGTGCCAG CAAGTCTTCGCCTCTGCGGAA
MI7K54	CCCCTGCCTTTGATTTCCAACCTTTCCCAAATGCAATACATGTGAAGCACACAGCACGAAACGCA CTCAGCAGGCTTTATCACTCCATAAAAAATAGGCACAACATTTACAAAAAATCAAAAAGAACTGG AAATTATCTGTAACAGTTTTGCAAATTCCTTCGAAGTAACTAAAAACA

Clone ID	GTSTs without the U3Cre-DNA junction
MI7K57	CTGAGGCAGTACAGGAGGGGTGAATTCCACCATGGAGAAAGTATTCAATAACAAGCCTACAGAGA AGTTGAGGAAGGTGGGAGGAGTAAAGGACAATGATGGTCCTCGAGAAACCTGCTCATGAAACCTG CTCAAGAAACCTGAGGAGGCAAAAGCTGATTCTAAAGTGTGAGAGAGCAACAGCACGAACCTGGG GAGGTGGGGTATCAAGTCCGAACCAAGGGTGAAGGGCAGGGACTCAGGTGCTGGAGTGTTTCAAT GTGGATCTTGGGGTTGCACTTGCTTGTTGAGTTACCTTGGGCAAGTTACTATGTGGTCTTAGTTT CTGTAAAATGGAGATCATGATGTAACCACTTGGTAAGGGCTGTTTGGAAGGATGGAAGAGTTCAG AAACAGAAAGCTCCCAGCACAAGCCT
MI7K59	GATTTAATCATATAATTTAATGAATCTGTTCTTTTTTTCCCCCCAAATATTTGTGCTTTAGGTGT AGGTTACCAGAATGATGAATTTTCCTCCTATGCTCCGTAGTCTTGTAATAAAAAGCATGTACAGT GTAGGACGTTTGCTAGGCATGGCTCTTCCTTTGACCAGTTCATGAACTGGAGTCAGGGTGGGGGT GAAGGCTAGGGTCCGAGGTCTGGGTCAGGCGCTCGTCACCGTCGTGGGAGGTCGAGGCTGATGTC CGCTTGCTCCCCATCAGCGTTAGTGCCCAGATCACCCCGTGCTCCGTCCCGCCGAAGCGGA
MI7K60	ACGCATAATGGGGGCTCTTGTTTCCGGGCTGAGCGTCATCTTTCTGTCTCCCTCTGGATTGAAAT GGCTTCTCAGCCCGAGAGG
MI7K63	CCACCACGACCTTTGCATAGGCCGTTCCCTCCGCTGGGCACAGCAATTCCCGCTGCTTTCTGCAT GGTTGCTCCTGAGACTCCTTCCCTCCCCC
MI7K64A	CATATTCACTAGGAGTTTAGGAGTTAGGAAGCCTAAGATAGAGAATATTACAGGCCCCACCATCC TATTGGTCATTATCGCCCCTCCACATATTCAAAACAGAATAGTCTCGTCTTCCTCAGACTTGCCT CCTTTATATTCCTGGCTTGGGGATGGGCATTGCTGTCTTTTCAGGGGTTCTTTGGGTCCTTCCTC GCCATTTT
MI7K65	CCTAAATTCTTTCAGCTAATTGATCTTCACTCGATTCGCGTTATCTATCTAGTATCACCATTCCG AGATGTCTTAAGTCTTAATCAGACAAGATAGTCTCTTTCGCAGAAGGCAGGAGTGCACTCCTTCC TTTCCTTGTATACGAGAAGGGGCAAAATCCACAGTTAAAGGCCTAGAAGCCGCGCACTCCG
MI7K69	AGTTCAGGCGTGAGCATAGAAAGCCAAAATATTTAGGCGGATAAACTCACGGAATATATTTT
MI7K71	CTCCCTGATTCACAAGAGAGCCTCCCTGGAAGCAAAGGCTGAGAGACAGCGCTCTCTCTCTCTCT TTCTGCATCTCCGGGAGA
MI7K72	GTAGGGCTCGTGGAAGGATGGAAGGAGTTCAGAAACAGAAAGCTCCCAGCACAGAGCCTGGCCCG TAATAAACAGTCTATGTGTTTGCTATCACGAGGGTACTTTTAAAGAAATTTAGTGCAGATGAGTA TAAATGAGACGTGGAGTGAGTAGCAGAATCAACACAAGGTTTTTCTTATAATGGATGAGAGATTG TAAAACACATTTGCATTGGGGAGAAGGACACATGGGAGAGGGAAGAAATGGAAGACACGGAAGAT TTGTTGGAGGCAAATTTAGATCCACGACAAAAGTACGGGAGTAAGTGCATGGGGAAAACTTAGTA TTGGAGGGAGGAAAAATGGCCTCACTTTGCACTTCGAATAACATCTACTTTCTTTCCTGGTCCTC CAAAGGCCACCAGGGCCAGCCTCTGGCCGTTCTTTGCCGTACTCCTCATTCTCCCCGGTGCATTT GCCATCTTGATTT
MI7K73	TTTGCCTGCCTTGGCCTACCAAAGTGCTAGAATTACAGGCATGAGCAACCATGATAACCAATATT CTTTATAATGAAAATAAATTAGAAAATGGTCAATATTTTCAGGTGGGAAAGTTAAAAAACTGTTT TTGGTTTTAAATT
MI7K74	CCCCCACCAGCCTATCCCTAACCTCTCATACCCTTCCCTCCTGAGGTCCTGACCTACAGAGAGCC CCTCTTGAAGGAGAGGGAGCCCCTGTAG
MII3K02	TAACGAACATAAAATACTTAGAACTTTGCCTGGGAGAAAATAAACAATATATGTGTTTAGTGTCA TCATTTTATTTTTTCTTTTCTTTCTGCTCTAATCATCACCGCTCCTGCCACCACGCAGCCATCAG GGAATATGAGAACTCATCTTCCTTCAATTCTTTTTTCCTGGATCTAACTTTGATTTAGCAGAATT GATTCTAACGTGCTATTCCCCAAGGCCACTCTGTTGCTTGTTCCTACCCAGAGTTACCCTTCCTT TGAAATCTGGTTGACAGTATGGCGGGCTTGGCAGGGAACCCGTCTTGCTCCACCCCGACGCCTGC TGCTGACCAGAGTTACCCTTGCTTTGAAATCTAGGTGACAGTGTGGG

Clone ID	GTSTs without the U3Cre-DNA junction
MII3K04	CCAGGCCTCCTGAGAGTACCCACCTGCCCGGCTTTCGGGGCTCATCTAAAAGGCTCCTCCGCAAG ACTAGGGGACACGGAATTGGGTTTGTTTAGCGGTCCTCTGAAGGGGCAAGATGGCGACAGGTAAC AGGGCGCAGGCTC
MII3K05	CCACAATTCAATTTTTGATGGCCAAACTGATTGCAATTTAATAACAACATGAGATACCATTTATT AGGAGCCACAGCACTTTTATTTTCTTTTTTTTTA
MII3K11	GGGCAGGTGGATCACCTGAGGTCAGGAGTTTGAGATCAACCTGACCAACATGATGAAACCCCATC TCTACTAAAAATACAAAAAATTAGCCGGGCATGGTGCAGACACCTGTAATCCTAGCTACTCATGA GCCTGAGGCAGGAGAATCACTTGAACTTGGGAGGCGGAGGTTGCAGTGAT
MII3K13	ATGTATAAAAACCCTCCAAAACATGGTGGTTTAAAACGATGCATTACGTTGATTACAATGATTCA TTCCTCTCCTGTGTCCTGGGGGTTGATGAACTCAGCTGGGTGGTTCTTGCTTATTGGGGTACCTC ATGTGGTTGCAGTCAGATAGTTGCTGATGCGGAGACATCTGAAGCTTCATCCCTT
MII3K27	GTTTCTCAACAGTCTAACTTCCCTTGGCTTTCCAGAGCCTTGATCTGCACTTGGATGGAAACCGT ATTCAGCTTGTAATTTTTCTTTGTAAACATTGCCTAATTGAAATTGCTCACTGGAACGGGGGTGT TGCATCTGAACAAAACCCTCTTGGACTTGGCCGCTGGGGTTACAGAGTAGTGAGTTGTGTTGAAA CTCCCCAGTAGAGCCAAAGGTCCTGTGACTTCCCAGAATCCACCAAGACTCTGGAAACAACCAGG AAAGTCTTTGAATGATTCACTCCAAAGCCCTGGCCCCATTGTTCATGGCCTCAAGCTCCTCCTGA GTATGGAGAAGACTGAGGAACACTTTGACAAAAGGGGGTTGGGTGTTTAAGTTATGACCCACCAG
MII3K34	AAAGCGAAGGGCGCCAAGGTAATGTGTGCCCTACACAGGGTCCTAGCCATAAGCACTTATGCAAG TCGGGGACAGGAGGAAAATTCTTGGTAACACTCCTCCCCTCAATTCTTGATCAGGGTCCCAGGCT CAACTTAATCAGTTCAGGGATGGGCCCCTTTGCACCTTTTTCC
MII3K35	GAGTTGAGATGCAAACACTTCCCAAGGAAGCAAGAATGGCCTCCTCCTCCACTCCCTAGAAGGAC CCAGGGCAACCGATCCACTTGACAACACCAGGGAATATGGGTTGCCAGCCACGCACTCATGAGAG AGGTGGCTTTGACTCCCAGACATGGATCCCAGGGGAGCCCAGGAACCTGTAGGAGTCGGTCTCCC AGTTGGCCACCATGGGATGGGCATTCGGGCCTGTGGCAGAGGG
MII3K36	CTGGTAGTCCAAAGAAACACGCACAGGGCCATAGATACAGCAACAGACTTACACAAACACAGAGG CCTCACCCAACCGCCCCCAGCCCTGGCTCCCCCCTCGCCTCCCCTACCCGCCCTCCAGCAGCCCT CCAGCAGCCGATCCCGGTCAAAGGGGAGGAGTCTTGGGCCCAGGAGCTTCGTCCCGGTGACAGGC CCGGCCAGACACCGGTGCACACTCCTGATGCTCCCTGCGGGGCTGCTGGAGGCTGGC
MII3K38	TGATGTATTCGAGCACGATTCTACTTAAGCCCTTTCCTTCCTGGATTTTGAGGGAGAATATCTTG CCTCTGTCCTTTAGGTTGACTGGAACATAGAGAACCCCAAAAGATCACGGAG
MII3K41	GAGACCTCCTGGCTAACATGGTGAAACCCCATCTCTACTAAAAATACAAAAAAAAAAATTTGCCG GG
MII3K46	CCCTTCTTTTTCCCAGCCCAGATATGACCTCTGAGAAAAGCTGAGGTCTTAGAGTCAAATCTGAG ACCCAGAATAGTTTGCAAGCTGTGTGACCTAGGGCTAGTGGCTGAACCTCTCTGGGCTTCCTGCT CCTTGATACTTACAAGGAAATGACACCACCCTTTCCCTCCCAGGTAATTTGGAAGGATTTCAATG CAGGAAAAGTGCCTGCCTTAT
MII3K48	TTAGAGAAAAAATACTGAAGCCTGGTGGAATCAGCACAAACTCTTAAGAGACTGGGGCTCAAATT CCAGCTCTACCAGCTGTAAGTTTAGCAGCAAGTTACTTAACATCTCAAAGTCTCGGTAAAAGACT CTGTAAAAAGAAGGTGATGCCTCTTTCAAAGTAGTTCTTAGTATTAAATTTTGAAAAAAGTAGGT AAATCTTAACCTCCTATCACACACGTAGTCTAATGACTTCCTAGGATCCTCCAACCCCAAGGGGA CCTCTTTCACAAAACCTCCCGGCTCCCCCCCA
MII3K55	CTAGACAGTGGTCTCAGCCTTTTTTTCTAATTTGATATATACCCGGAACAAATGACACTCACTTG CCAGAATGTCTCCACCCTACAGCTTCTCAAACTGTGTGATGTAATATATTGACCCCCCAGTCCTC AGGGCCTTTGGGTAGGATGTTAGGGTGGGTGACTAGAGCCCTTGG

Clone ID	GTSTs without the U3Cre-DNA junction
MII3K65	GTGGCCTGCTCAGTCTTTTATCTGTTCAATGGGTCCGGCGCTCCCACTCAGCCCTCGGACAGCCC CGGTGTCTGCGCCGTCCCAGGTTGTCTCCGCGCCTCCCGTTCCGCCGGCGTTCCCTGGCCTCCGC GCGC
MII3K74	AGACCTCCTGGCTAACATGGTGAAACCCCTCTCTACTAAAAATACAAAAAAAAAA
MIK78	CTGTCCCTGGATTCCAAGGACACGCCTCCCTGGGAAGCAAAGGCTGAGAGACAGCGGCTCTCTCT CTCTCTTTCTGCATCT
MI7K77	CCACCACGAGATCCTTTGCATAGGCCGTTCCCTCCGCTGGGCACAGCATTCCCGCTGCTTTCTGC ATGGTTGCTCCTGAGACTCCTTCCCTCCCCCGA
MIK83	AAGCAAGGGCGCCCCAAGGTAATGTGTGCCCTACACAGGGCTCCTAGCCATAAGCACTTATGCAA GTGCGGGGCACAGAGAAAATTCTGGTAACACTCCTCCCCTCAATTCTGGATCAGGTCCCAGGCTC AACTAATCAGTTCAGGGATGGGCCCTTTGCCACCTTTTTCCT

9.4. Bioinformatics analysis

Homology of the inverse PCR sequences with the human genome recovered different genes which are listed in the following table (table 7).

Clone ID	Enzyme for iPCR	Gene (GeneCard nr.)	Transcript (Ensembl/NCBI)	Direction of integration	Location within gene
MI3K01	*Pst*I	SLC12A2 (GC05P127447)	ENST00000262461 NM_001046.2	antisense	1013 bp upstream
MI3K01	*Pst*I	–	ENSESTT000000 48701	sense	exon 1
MI3K02	*Pst*I		RP11-10N16.3-001	antisense	intron 1
MI3K02	*Pst*I	GRHL3 (GC01P024392)	ENST00000361548 NM_198173	sense	intron 1
MI3K03 MI3K62 MI3K11 (identical)	*Pst*I	S100A10 (GC01M148768)	ENST00000358003 NM_002966.1	antisense	intron 1
MI3K03 MI3K62 MI3K11 (identical)	*Pst*I	Genscan	–	sense	intron 2
MI3K04	*Nsp*I	HKII (GC02P074971)	ENST00000290573 NM_000189.4	sense	intron 02

Clone ID	Enzyme for iPCR	Gene (GeneCard nr.)	Transcript (Ensembl/NCBI)	Direction of integration	Location within gene
MI3K05	*Pst*I	TOP1 (GC20P039090)	ENST00000361337 NM_003286.2	antisense	565 bp upstream
MI3K09	*Pst*I	USP40 (GC02M234168)	ENST00000251722 NM_018218.1	antisense	intron 4
MI3K10	*Pst*I	JUNB (GC19P012763)	ENST00000302754 NM_002229.2	sense	1950 bp upstream
MI3K12	*Pst*I	PRKCBP1 (GC20M045271)	ENST00000311275 NM_183047.1	antisense	2206 bp upstream
MI3K13	*Pst*I	PITPNM2	ENST00000280562	antisense	exon 26
MI3K13	*Pst*I	Q8TEM4_ HUMAN	ENST00000300798	sense	2016 bp downstream
MI3K14	*Pst*I	Repetitive sequence			
MI3K16	*Pst*I	CT142/ C20orf142 (GC20M042364)	ENST00000360847 XM_371399.2	sense	intron 1
MI3K17A	*Pst*I	HNRPH1 (GC05M178973)	ENST00000356731 NM_005520.1	sense	exon 1
MI3K18	*Pst*I	Genscan	–	sense	intron 1
MI3K21	*Pst*I	SPATA 5	ENST00000274008 NM_145207	sense	intron 14
MI3K22	*Nsp*I	Genscan	–	antisense	exon 1
MI3K22	*Nsp*I	Genscan	–	sense	downstream
MI3K23	*Nsp*I	Genscan	–	antisense	intron 1
MI3K23	*Nsp*I	LIPC (GC15P056511)	ENST00000299022 NM_000236.1	sense	intron 1
MI3K24	*Pst*I	Repetitive sequence			
MI3K36 MI7K27B (identical)	*Nsp*I	Genomic sequence			–
MI3K37	*Nsp*I	PRKCBP1 (GC20M045271)	ENST00000311275 NM_183047.1	antisense	1584 bp upstream

Clone ID	Enzyme for iPCR	Gene (GeneCard nr.)	Transcript (Ensembl/NCBI)	Direction of integration	Location within gene
MI3K39A	*Nsp*I	EST	ENSESTT000000 69582	sense	intron 2
MI3K40	*Nsp*I	ABCC3 (GC17P046067)	ENST00000285238 NM_003786.2	sense	intron 8
MI3K44 MI7K59 (identical)	*Pst*I	KCTD5 (GC16P002672)	ENST00000301738 NM_018992.1	antisense	428 bp upstream
MI3K44 MI7K59 (identical)	*Pst*I	P461 (GC16M002592)	ENST00000356048 AC005591	sense	2198 bp upstream
MI3K47	*Pst*I	Genscan	–	sense	201 bp upstream
MI3K61	*Pst*I	PRSS36	ENST00000268281 NM_173502.2	sense	147 bp downstream
MI3K66	*Pst*I	RNF184 (GC03M137351)	ENST00000309993 NM_018133.2	antisense	1085 bp upstream
MI3K67	*Pst*I	FGD3 (GC16M002592)	ENST00000262555 AL389924	antisense	intron 2
MI3K76	*Pst*I	Genscan	–	antisense	intron 2
MI3K76	*Pst*I	CTNND2 (GC05M011024)	ENST00000304623 NM_001332.2	sense	intron 1
MI3K80 MI7K57 (identical)	*Pst*I	ESR1-005 (GC06P152220)	ENST00000206249 NM_000125	antisense	intron 1
MI3K81 MI7K72 (identical)	*Pst*I	ESR1-005 (GC06P152220)	ENST00000206249 NM_000125	antisense	intron 1
MI7K07	*Pst*I	Repetitive sequence			
MI7K08	*Nsp*I	Repetitive sequence			
MI7K20	*Pst*I	KIF11	ENST00000260731 NM_004523.2	sense	1057 bp upstream
MI7K20	*Pst*I	Genscan	–	sense	intron 3

Clone ID	Enzyme for iPCR	Gene (GeneCard nr.)	Transcript (Ensembl/NCBI)	Direction of integration	Location within gene
MI7K26	*Pst*I	Genscan	–	antisense	intron 1
MI7K28	*Pst*I	SLC7A2 (GC08P017440)	ENST00000004531 NM_003046	antisense	21039 bp upstream
MI7K28	*Pst*I	Genscan	–	sense	3966 bp downstream
MI7K29	*Nsp*I	EED (GC11M118272)	ENST00000351625 NM_003797.2	antisense	exon 1
MI7K30B	*Nsp*I	ZNF143 (GC11P009439)	ENST00000299606 NM_003442.3	antisense	654 bp upstream
MI7K31	*Nsp*I	Genomic sequence	–	–	–
MI7K32	*Pst*I	Repetitive sequence			
MI7K34	*Nsp*I	Genscan	–	sense	intron 4
MI7K48	*Pst*I	ARHGAP11A (GC15P030694)	ENST00000361235 NM_014783.2	antisense	813 bp upstream
MI7K48	*Pst*I	EST	ENSESTT000000 49155	sense	exon 1
MI7K49	*Nsp*I	PARD6B (GC20P048781)	ENST00000262600 NM_032521.1	antisense	29517 bp upstream
MI7K50	*Pst*I	Several ESTs	–	sense	intron 1
MI7K51	*Pst*I	PRKCBP-1 (GC20M045271)	ENST00000262975 NM_183047	antisense	intron 2
MI7K51	*Pst*I	Q96N05_human	ENST00000360965	sense	70 bp downstream
MI7K53	*Nsp*I	ZNHIT2 (GC11M064640)	ENST00000310597 NM_014205.2	sense	exon 1
MI7K54	*Pst*I	YARS (GC01M032909)	ENST00000257116 NM_003680.2	antisense	564 bp upstream
MI7K54	*Pst*I	S100PBPR (GC01P032953)	ENST00000263536 NM_022753.2	sense	intron 1
MI7K60	*Pst*I	Genscan			

Clone ID	Enzyme for iPCR	Gene (GeneCard nr.)	Transcript (Ensembl/NCBI)	Direction of integration	Location within gene
MI7K63 MI7K77 (identical)	*Pst*I	Genomic sequence	–	–	–
MI7K64A	*Pst*I	EST	ENST00000362134	–	–
MI7K65	*Pst*I	Genomic sequence	–	–	–
MI7K69	*Pst*I	Genomic sequence	–	–	–
MI7K73	*Nsp*I	Genomic sequence	–	–	–
MI7K74	*Pst*I	ZBT7B/ZFP67 (GC01P151790)	ENST00000326770 NM_015872.1	antisense	675 bp upstream
MII3K02	*Pst*I	C9orf3/FLJ14675 (GC09P094568)	ENST00000311678 NM_032823.3	antisense	intron 5
MII3K04	*Pst*I	Genscan	–	antisense	121 bp upstream
MII3K04	*Pst*I	EMSY (GC11P075833)	ENST00000343878 NM_020193.3	sense	intron 1
MII3K05	*Pst*I	ZNRF1 (GC16P073590)	ENST00000320619 AL834440	sense	intron 1
MII3K11	*Pst*I	NTN4 (GC12M094554)	ENST00000343702 NM_021229.2	antisense	10643 bp upstream
MII3K11	*Pst*I	EST	ENSESTT0000009 4509	sense	1845 bp upstream
MII3K13	*Nsp*I	Genscan	–	sense	6769 bp? downstream
MII3K27	*Pst*I	SLC1A2 (GC11M035229)	ENST00000278379 NM_004171.2	sense	intron 5
MII3K34	*Pst*I	Genomic sequence	–	–	–
MII3K35	*Nsp*I	TFF1 (GC21M042655)	ENST0000291527 NM_003225.2	antisense	intron 1
MII3K35	*Nsp*I	EST	ENST00000339657	sense	intron 1
MII3K38	*Pst*I	Genscan	–	sense	exon 14
MII3K41	*Pst*I	Repetitive sequence	–	–	–

Clone ID	Enzyme for iPCR	Gene (GeneCard nr.)	Transcript (Ensembl/NCBI)	Direction of integration	Location within gene
MII3K46	*Pst*I	C1QTNF6 (GC22M035900)	ENST00000255836 NM_182486.1	sense	intron 1
MII3K48	*Pst*I	WDR10 (GC03P130641)	ENST00000348417	antisense	904 bp upstream
MII3K48	*Pst*I	MBD4 (GC03M130632)	ENST00000249910 NM_003925.1	sense	intron 1
MII3K55	*Pst*I	TEX14 (GC17M053989)	ENST00000240361 NM_198393.2	antisense	intron 1
MII3K65 MII3K36	*Pst*I	BCL9L (GC11M118272)	ENST00000334801 NM_182557.1	antisense	156 bp upstream
MII3K71	*Pst*I	PRKCBP1 (GC20M045271)	ENST00000311275 NM_183047.1	antisense	2641 bp upstream
MIK78 MI7K71 (identical)	*Pst*I	PRKCBP1 (GC20M045271)	ENST00000311275 NM_183047.1	antisense	2215 bp upstream
MIK83	*Nsp*I	Genomic sequence	–	–	–

Table 7. Bioinformatic analysis of gene trap sequence taqs (GTSTs). Clone ID denotes the cell clone obtained after the final selection for cells with gene trap integrations in TNFα inducible loci, enzyme for inverse PCR indicates the enzyme used in the inverse PCR, if *Pst*I was used the retrieved GTST is upstream to the U3Cre integration site, if *Nsp*I was used, downstream. Gene indicates the gene in or close to which the gene trap integration occurred together with the unique gene card identifier (www.genecards.org). Other classification categories are EST (homology to expressed sequence tag); GenScan, when only a gene prediction was identified; repetitive sequence and genomic sequence, when the integration occurred in a region, where no genes have been assigned or predicted. Transcript provides the transcript accession number of the recovered gene from the Ensembl and NCBI databases. Direction of integration and location within gene indicate the orientation of the gene trap encoded Cre gene relative to the recovered gene and the position of the U3Cre gene trap within this gene.

Curriculum Vitae

Personal particulars

Name:	Francisca Guardiola Serrano
Address:	Wittelsbacherallee 77, 60385 Frankfurt am Main, Germany
Marital status:	Single
Nationality:	Spanish
Date of birth:	30-05-1972
Place of birth:	Palma de Mallorca

Education

Degree

- Graduate in Sciences (Biology). *Balearic Islands University*, Palma de Mallorca (from 1990 to 1995). June 1995.

- Graduate in Biochemistry. *Balearic Islands University*, Palma de Mallorca (from 1995 to 1997). June 1997.

- Since 2001 Ph.D. student, *University Hospital Frankfurt am Main*, department of Molecular Hematology.

Other University education

- University specialist in Food Analysis. Postgraduate course from *Pla Mestral* of 220 theoretical-practical hours, founded by *Fondo Social Europeo*, *Balearic Island University*, Palma de Mallorca, 1997.

- University specialist in Molecular Biomedical techniques. Postgraduate course from *Pla Mestral* of 205 theoretical-practical hours; founded by *Fondo Social Europeo*, *Balearic Island University*, Palma de Mallorca, 2000.

Experience

- Assist in practical training for *Methods and experimental techniques in Molecular and Cellular Biology* subject of the Biology and Biochemistry degree at the *Balearic Islands University*, academic year 96-97.

- Practical training, *German Institute of Nutrition* (Bergholz-Rehbrücke) from October 1997 to January 1999, with a studentship founded by *Sa Nostra*.

- Teaching assistant of 4 hours in the *Balearic Island University, Department of Basic Biology and Health Sciences* (February 2000- December 2000).

- Ph.D. student at the University Hospital Frankfurt am Main working on the project *Identification of survival genes in breast carcinoma* in the department of Molecular Hematology of Prof. Dr. Harald von Melchner under supervision of Dr. Joachim Altschmied (August 2001- December 2005).

Coursework

- *Metabolical and pathological nutrients Interrelationships*. Course of the Ph.D. program in *Molecular Nutrition*, *Balearic Island University*, academic year 97-98.

- *Molecular and Biochemical Biology approximation to the body weight control*. Course of the Ph.D. program in *Molecular Nutrition*, *Balearic Island University*, academic year 97-98.

- *Molecular and nutritional bases of cancer*. Course of the Ph.D. program in *Molecular Nutrition*, *Balearic Island University*, academic year 98-99.

- *Biochemical and molecular biology from brown adipose tissue*. Course of the Ph.D. program in *Molecular Nutrition*, *Balearic Island University*, academic year 97-98.

- *Molecular Nutrition, technological possibilities*. Course of the Ph.D. program in *Molecular Nutrition*, *Balearic Island University* academic year 97-98.

- *Molecular Biology techniques*. Course of the Ph.D. program in *Molecular Nutrition*, *Balearic Island University*, academic year 98-99.

- *Perinatal nutrition*. Course of the Ph.D. program in *Molecular Nutrition*, *Balearic Island University*, academic year 98-99.

- Pedagogic adaptation course (CAP, 2000-2001).

Published abstracts

Gianotti, M., Bonet, M., **Guardiola F**., Palou A. Application of the direct isothermal microcalorimetry to the study of thermogenesis in isolate adipocytes.
National Congress from the Spanish Society of Basic and Applied Nutrition, Madrid 20-23.11.1996

Altschmied J., **Guardiola-Serrano F**., Strolz C., Sturm K. and von Melchner, H. (2002). Isolation of genes transiently induced by TNFalpha.
The Genetic and Molecular Basis of Human Disease. Symposium of the National Genome Network (NGFN) and the German Human Genome Project (DHGP), Berlin 17.11.-19.11.2002

Guardiola-Serrano F., von Melchner, H. and Altschmied J. (2003). Characterization of TNFalpha induced antisense transcripts in a mammary carcinoma cell line.
Genes, Chromosomes, Genomes 9, 76

Altschmied J., **Guardiola-Serrano F**., Sturm K. and von Melchner, H. (2003). Trapping of TNFalpha induced genes reveals a high proportion of antisense transcripts.
Genes, Chromosomes, Genomes 9, 38

Altschmied J., **Guardiola-Serrano F**., Sturm K. and von Melchner, H. (2004). TNFalpha signalling: Gene regulation by antisense transcripts?.
Annual Meeting of the German Society for Cell Biology, Berlin 24.03.-27.03.2004

Altschmied J., **Guardiola-Serrano F**., Sturm K., Strolz C. and von Melchner, H. (2004). Characterization of survival genes recovered from a functional gene trap screen in human cancer cell lines.
NGFN-Meeting, Berlin 20.11.-21.11.2004

Guardiola-Serrano F., von Melchner, H. und Altschmied J. (2005). Characterization of TNFalpha induced survival genes in a mammary carcinoma cell line.
Annual Meeting of the German Society for Cell Biology, Heidelberg 16.03.-19.03.2005

Guardiola-Serrano F., Corinna Strolz, Harald von Melchner and Joachim Altschmied. Identification of apoptosis regulators with a gene trap approach.
Molecular approaches to controlling cancer, LXX Cold Spring Harbor Symposium, Cold Spring Harbor 1.06 -6.06.2005

Scholarships

(October 97- January 99). Scholarship "*Beca para la ampliación de estudios en el extranjero"* sponsored by "*Sa Nostra. Caixa de Balears*", German Institute of Nutrition, Bergholz-Rehbrücke.

GlaxoSmithKline Stiftung Reisekostenbeihilfe, CSHL Symposium, Cold Spring Harbor, USA, 1-6 June 2005.

Ehrenwörtliche Erklärung

Ich erkläre hiermit ehrenwörtlich, daß ich die dem Fachbereich Biologie zur Promotionsprüfung eingereichte Arbeit mit dem Titel

„Identification and characterization of TNFα responsive genes in human breast cancer cells“

im Zentrum der Inneren Medizin, Medizinische Klinik II, Abteilung Hämatologie des Universitätsklinikums Frankfurt bei Prof. Dr. Harald von Melchner und unter Leitung von Prof. Dr. Anna Starzinski-Powitz mit Unterstützung von Dr. Joachim Altschmied ohne sonstige Hilfe selbst durchgeführt und bei der Abfassung der Arbeit keine anderen als die in der Dissertation angeführten Hilfsmittel benutzt habe.

Ich habe bisher an keiner in- oder ausländischen Biologischen Fakultät ein Gesuch um Zulassung zur Promotion eingereicht noch die vorliegende Arbeit als Dissertation vorgelegt.

Frankfurt, den 26.06.2006

Dipl. Biol. Francisca Guardiola Serrano