Aus der Abteilung Neurogenetik

(Direktor: Prof. K.-A. Nave, Ph. D.)

Arbeitsgruppe Genexpression Prof. Dr. rer. nat. M. J. Rossner

des Max-Planck-Instituts für Experimentelle Medizin

in Göttingen

Comprehensive analysis of transcription factor activity monitoring with *Cis*-elements coupled EXTassys in living cells

INAUGURAL – DISSERTATION

zur Erlangung des Doktorgrades

der Medizinischen Fakultät der

Georg-August-Universität zu Göttingen

vorgelegt von

Anna-Katharina König

aus

Rauenberg

Göttingen 2017

Dekan:	Prof. Dr. rer. nat. H. K. Kroemer
Referent:	Prof. Dr. rer. nat. K.-A. Nave
Ko-Referent/ in:	Prof. Dr. Steven Johnsen
Drittreferent/ in:	Prof. Dr. Dieter Kube

Datum der mündlichen Prüfung: 04.07.2018

Hiermit erkläre ich, die Dissertation mit dem Titel ‚‘Comprehensive analysis of transcription factor activity monitoring with *Cis*-elements coupled EXTassys in living cells‘ eigenständig angefertigt und keine anderen als die von mir angegebenen Quellen und Hilfsmittel verwendet zu haben.

Göttingen, den 12.11.2017 .....

(Unterschrift)

Table of contents

List of figures and tables

Figures:

Tables:

Abbreviations

ATCC	American Tissue Culture Collection
ATP	Adenosine TriPhosphate
cDNA	complementary DNA
Cis-	a prefix meaning ‘on the same side’
CLS	Cell Lines Service GmbH
CMV	CytoMegaloVirus
Co-IP	Co-Immunoprecipitation
dbcAMP	Dibutyryl-cyclic 3‘5‘-Adenosine MonoPhosphate
DMEM	Dulbeco’s Modified Eagle Medium
DMSO	DiMethylSulphOxide
DNA	DeoxyriboNucleic Acid
DNAse	Deoxyribonuclease
dNTP	DeoxyriboNucleotide TriPhosphate
DTT	1,4- DiThioThreitol
ECACC	European Collection of Cell Cultures
ECL	Enhanced ChemiLuminescence
E. coli	*Escherichia coli*
EDTA	EthyleneDiaminTetraAcetate
EGTA	EthyleneGlycolTetraAcetate
et al.	and others
EtBr	Ethidium Bromide
EtOH	Ethanol
EXT	*Ex*pressed unique sequence *T*ag
EYFPnuc	nuclear localized Enhanced Yellow Fluorescent Protein
FBS	Fetal Bovine Serum
f.c.	final concentration
HS	Horse Serum

LB	Luria-Bertani
LDS	Lithium Dodecyl Sulfate
MEM	Minimum Essential Medium
MOI	Multiplicity of Infection
mRNA	messenger-RNA
NCBI	National Center for Biotechnology Information
PBS	Phosphate-Buffered Saline
PCR	Polymerase Chain Reaction
pH	negative decimal logarithm of the hydrogen ions (H^+)
PLL	poly-L-lysine
PMA	Phorbol Myristate Acetate
RLUs	Relative Luciferase Units
RNA	RiboNucleic Acid
RNAse	Ribonuclease
SD	Standard Deviation
SDS	Sodium Dodecyl Sulfate
SOC	Super Optimal Broth medium with Catabolite repression
TAE	Tris-Acetate-EDTA
Taq	Thermus aquaticus
TE	Tris/EDTA
TK	Thymidinkinase promoter (from herpes simplex virus)
Tm	Melting temperature
Tris	Tris(hydroxymethyl)aminomethane
UV	Ultraviolet

DNA and RNA bases:

A	adenine
C	cytosine
G	guanine
N	any base

S	G or C
T	thymine
U	uracil
W	A or T

Units of measurement:

A	absorbance
b	bases
bp	basepairs
°C	degree Celsius
g	gram
kb	kilobase
L	liter
M	molar
min	minutes
n	number
OD	optical density
rpm	rotations per minute
sec	seconds
U	units of enzymatic activity
V	volume
v/v	volume/volume
w/v	weight/volume

Power prefixes:

m	milli (10^{-3})
µ	micro (10^{-6})
n	nano (10^{-9})
p	pico (10^{-12})
f	femto (10^{-15})

1. Introduction

1.1 Regulatory elements of transcription

Cis-elements

Cis-elements or *Cis*-acting DNA sequences are gene regulatory elements that can be recognized by transcription factors. *Cis*-elements are non-coding DNA regions. Two groups of *Cis*-elements exist in mammals: the long-range regulatory elements and the proximal promoter elements.
Long-range regulatory elements are DNA sequences that function most likely as enhancer or silencer regions. These are typically -1000 to -700 bp or more distant from the start of transcription and have a length of approximately 500 bp. They contain up to ten transcription factor binding sites within one region. Enhancers increase the gene promoter activity which leads to an increase of transcription. Long-range regulatory elements leading to a repression of the gene promoter activity are called silencers. There exist other long-range regulatory elements than enhancers and silencers, such as insulators, locus control regions and matrix attachments regions.

Proximal promoter elements, sometimes called upstream regulatory elements, are most likely located in clusters and serve as transcription factor binding sites. Their location in promoter regions is just 5' to the core promoter elements and near the transcription start site. If these *Cis*-elements are located near the transcription start site, they can increase the transcription rate of the gene located 3' of the promoter.

The transcription rate per gene is regulated by the entity of all *Cis*-elements, of enhancers and silencers.

Core promoter elements

The core promoter plays a pivotal role in the regulation of transcription, especially in its initiation (Smale and Kadonaga 2003). The core promoter is found approximately 35 bp up- and/or downstream from the transcription start site (+1). Within this promoter region the elements themselves can interact directly with parts of the basal transcription machinery. This group of minimal essential proteins for transcription includes the RNA polymerase II itself and other transcription factors as e.g. TFIID/TFIIA or TFIIB. Possibly the most important and certainly the best known core promoter elements are the TATA box, the initiator element, the BRE (TFIIB recognition element) and the downstream promoter element. These elements are found only in a subset of all mammalian gene promoters. A core promoter may contain a single, a group, or none of these elements. Some of the elements can work autonomously, others have to be combined for functionality. The well-known TATA box with its consensus sequence TATA(A/T)AA(G/A) can function without any other core promoter element and is the binding site for the TATA-binding protein, a major subunit of the TFIID. However, it is found in only in 32% of all human gene promoter regions.

Transcription factors

Transcription factors or *trans*-acting factors are proteins with the ability to bind to and interact with specific DNA sequences called *Cis*-elements. Via this DNA interaction, these proteins are able to regulate the gene activity at the level of transcription. The protein amounts of transcription factors themselves depend on the transcriptional regulation of their genes, activation or repression via proteolysis, ligand binding or modifications like

phosphorylation. Transcription factors (TFs) can act as activators and/or repressors of the transcription depending on the *Cis*-elements - their specific binding site or an interaction with other proteins. A repressor leads to downregulation of the transcription via blocking the general machinery, whereas activators increase the transcription rate. Transcription factors can be subdivided in superclasses, classes, families and subfamilies. The three major superclasses contain the zinc finger transcription factors, the helix-turn-helix domain transcription factors with the homeodomain as a class, and the basic domain transcription factors with the basic leucine zipper class and the basic helix-loop-helix class. Minor superclasses are the all-α-helical DNA-binding domains, α-helices exposed by β-structures, Immunoglobulin fold, β-hairpin exposed by a α/β scaffold, β-sheet binding to DNA, β-barrel DNA-binding domains, and the as yet undefined DNA-binding domains (Wingender et al. 2013).

1.2 Reporter gene assays

Reporter gene assays are wide-range measurement tools in molecular biology, biochemistry and pharmaceutical research. They have a broad range of applications and are often used to assess activity of regulatory elements (promoter or *Cis*-element), to measure transcription factor activity, to assess gene expression or as a selection marker. A reporter gene has two functional parts. The reporter that encodes for a protein and the *Cis*-regulatory element (or promoter) that drives the transcription. The reporter protein itself can easily be detected and quantified. The most commonly used reporter proteins are enzymes, fluorescent proteins or selection markers as antibiotic resistance (Bronstein et al. 1994).
Some of the reporter gene assays are used for highly sensitive quantitative analysis, such as enzyme encoding reporter genes. This enzymatic activity is measured by chemiluminescence, fluorescence or light absorption, depending on the substrate used. Each enzyme molecule transforms substrate molecules corresponding to the enzymatic nature. This enzymatic readout leads to signal amplification. The most common reporter enzymes used for these readouts are different luciferases, β-galactosidase and β-lactamase. In pharmacological research, gene reporter assays are used for drug discovery in high throughput screenings. These screenings are helpful to find new receptor ligands out of chemical libraries. Reporter gene assays play an important role in the analysis of signal transduction and transcriptional regulation.

Specific binding sites (*Cis*-elements) for TFs are often located near 5'to a gene or in a promoter of a gene. The activity of specific transcription activating TFs can be measured via gene reporter assays with isolated *Cis*-elements as promoter region (Levine and Tjian 2003). Protein/DNA binding for regulating but not directly acting as TFs can be measured by one hybrid method (Deplancke et al. 2004). protein/protein interactions in the nucleus can be assessed by two hybrid methods (Luo et al. 1997). Protein/protein interactions at plasma membranes and in the cytosol can be measured by use of multiple protein complementation assays that are coupled to the transcriptional readout (Stagljar et al. 1998). Applying these assays, the main body of the signal transduction network can be explored. Today, reporter gene assays are essential and flexible measurement tools with a broad range of applications that can detect a wide variety of cellular events by selection of assay design and the appropriate *Cis*-elements.

1.3 Multiplexed reporter gene assays

Transcription factor activity measurement through classical reporter gene assays is time- and cost-consuming. In a classical reporter gene assay, the activity of only one TF can be

monitored. This reporter construct has a specific *Cis*-element coupled to a minimal core promoter element and a reporter gene like β-galactosidase, luciferase or another enzymatic reporter protein. The activated TF recognizes the *Cis*-element, binds to it and initiates the transcription of the reporter gene. In classical reporter gene assays, the generation of data is based on the expression of proteins. For multiplexing reporter gene assays the readout has to be on the transcriptional level and not on translational level. Multiplexing can be achieved using a nucleic-acid based reporter by adding one unique restriction site at different locations resulting in homogeneous cDNA reporters with different but defined sizes. Using a capillary electrophoresis, the reporters can be separated by size and analyzed (Romanov et al. 2008). This approach reduces the background signal and yields robust and sustained cell signatures. However, the readout was done with only 43 reporter constructs and upscaling is limited to a few hundreds. For comprehensive genome-wide TF activity analysis this method cannot be used. Another method is to use reporters based on nucleic acids using unique expressed oligonucleotides that perform as RNA barcodes can be multiplexed up to high numbers. Analyzing the expression of a used nucleic acid reporter library by microarray technology or next generation sequencing the activity of the corresponding TFs can be measured (Li et al. 2006). Theoretically, this method can be scaled up without limit and a high number of TF activities can be measured simultaneously.

Multiplexed reporter gene assays are very useful for genome-wide analysis of transcription factor activity and signal transduction. Further development of methods for reporter gene assay multiplexing at a large scale analysis is needed.

1.4 High-throughput technologies

The generation of datasets in molecular biology research using e.g. classical reporter gene assays is extremely time- and cost- consuming. To test e.g. the activity of one TF one reporter gene construct with the corresponding *Cis*-element have to be cloned and tested within cell culture experiments under different conditions. In TF activity screenings e.g. high-throughput technologies enable the measurement of many different TFs and/or several conditions within the same time.

The vast development in computational science enabled a revolutionary process in the whole –OMICS field (e.g. genomics, transcriptomics, proteomics etc.). Automated equipment allows an upscaling of experiments without influencing their quality. High-throughput screenings in drug discovery are often performed with the support of robotics. High throughput technologies are not only used in drug development and analysis of cell signalling (Chanda and Caldwell 2003). Also DNA and RNA sequencing via NGS (Next Generation Sequencing), microscopy and imaging technologies or flow cytometry are common fields or methods.

1.5 EXTassays

EXTassays represent a novel technique to monitor cellular signaling within living cells. It is a highly scalable reporter system using expressed oligonucleotide tag (EXT) as a nucleotide reporter instead of classical reporter proteins (Botvinnik et al. 2010). In comparison to classical reporter systems based on reporter proteins, EXTs perform better in kinetics and sensitivity. The EXT library performs with balanced melting temperature and virtually no intramolecular complementary regions. Each EXT is an oligonucleotide with its length of 49 bases and consists of a core region flanked of 10 ‘words’. ‘Words’ are sequence stretches of 4 nucleotides consisting of one cytosine and three adenosines or thymidines. A core region is a variable region with a length of 9 nucleotides. Three central nucleotides consisting of

cytosine and guanine are flanked by alternating adenosine, thymidine, cytosine and guanine (Botvinnik et al. 2010). An EXT is flanked by 5' and 3' with invariable primer regions for amplification and cloning.

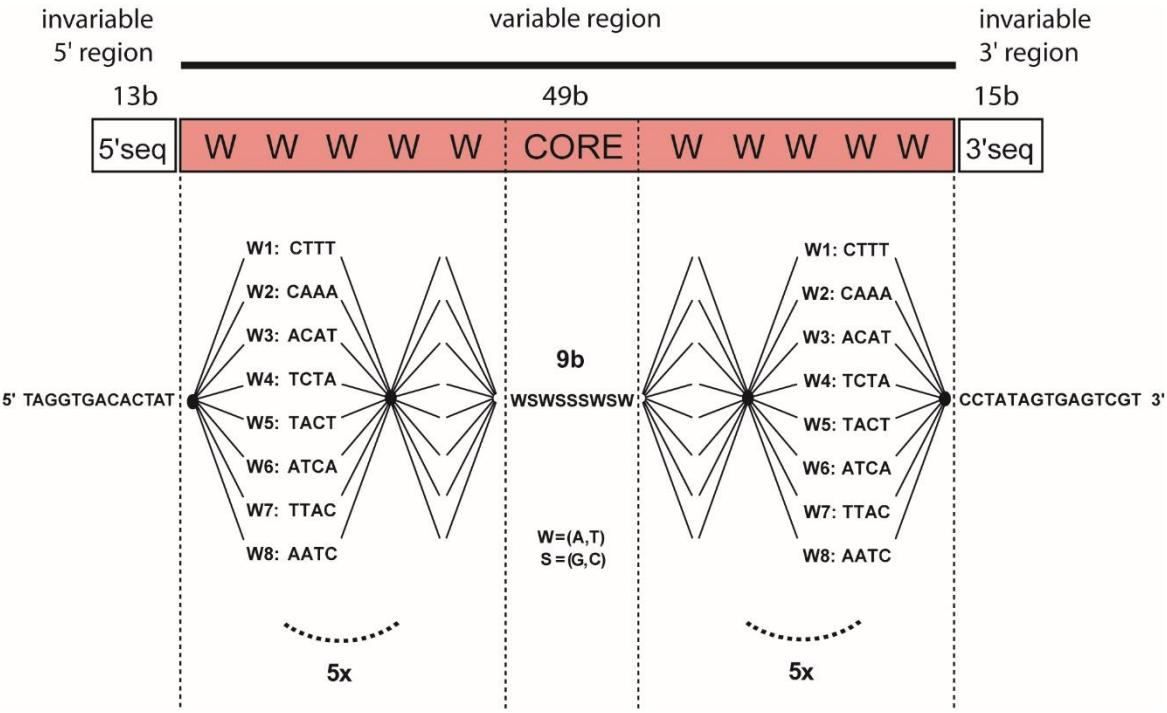


Figure 1.1: design of the EXT

The core region consisting of 9 bases is 5' and 3' flanked of 5 words. Eight different words exist and consist of four bases. Every EXT is an oligonucleotide of 49 bases and is 5' and 3' flanked on primer sequences.

1.6 Cancer

The World Health Organization (WHO) describes cancer as follows: "Cancer is the uncontrolled growth and spread of cells. It can affect almost any part of the body. The growths often invade surrounding tissue and can metastasize to distant sites. Many cancers can be prevented by avoiding exposure to common risk factors, such as tobacco smoke. In addition, a significant proportion of cancers can be cured, by surgery, radiotherapy or chemotherapy, especially if they are detected early."(http://www.who.int/topics/cancer/en/).

Cancer comprises a large group of subtypes that can affect almost every organ or cell type in the human or animal body. Synonyms for cancer are neoplasm, malignant tumor or malignant process. The terminology of cancer depends on the tissue or cell origin. Cancer originating from epithelial cells is called carcinoma. A subgroup the adenocarcinomas are derived from epithelial glandular cells. Sarcomas are cancers derived from mesodermal cells e.g. bone, muscle, vessels, fatty- or connective tissue cells. Cancer from blood cells, especially granulocytes, monocytes, or lymphocytes, is called leukaemia or lymphoma.

Cancer is a genetic disease driven by mutations in the genome. The majority of alterations in the genome are somatic mutations. A minority of cancers are hereditary origin with mutations in the germline genome. Syndromes as Li-Fraumeni with a germline mutation in *TP53* or HNPCC (hereditary non-polyposis colorectal cancer) or Lynch-syndrome with microsatellite

instability take a higher incidence of carcinogenesis of divers cancers compared to the normal population.

In some cancer types a so called adenoma-carcinoma sequence is known. In the model of adenocarcinomas of the colon the first driver mutation is the loss of *APC*, that cause the formation of an adenoma in the epithelial cell layer. Several following mutations in these adenoma cells containing *KRAS*, *DCC* and *TP53* lead to the arising of an invasive adenocarcinoma. Also in a subgroup of the pancreatic ductal adenocarcinomas (PDACs) an adenoma-carcinoma sequence exist. The intraductal papillary mucinous neoplasia (IPMN) show a progression that follows an adenoma-carcinoma sequence including mutations in *GNAS*, *KRAS* and *TP53*.

Common driver mutations in most cancers are loss of function mutations or deletions that occur in tumour suppressor genes (e.g. *TP53*, *CDKN2A*) and DNA repair mechanism genes (e.g. *BRCA1*, *BRCA2*). Gain of function mutations or activating mutations often occur in genes coding for receptors or kinases (e.g. *EGFR*, *KRAS*) that are involved in cell proliferation processes.

Malignant tumours are characterized with the nature of prolonged viability and resistance to cell death stimuli, enhanced proliferative signalling and promotion of neo-angiogenesis to support optimal nutrition of the tumour. Also mutations in genes coding for cell skeletal proteins and cell-cell adhesion molecules play a pivotal role leading to enhanced invasive growth and metastasis. Not only enhanced growth and invasion are typical characteristics of cancer, also modulation of the immune system with tumour promoting inflammation and tumour masking to avoid tumour cell destruction via the immune system are central features (Hanahan and Weinberg 2011).

Cancer types with the highest incidences in the western civilization are lung, colorectal, prostate and breast cancer. Cancer diseases with the worst 5 year survival rate are small cell lung cancer (SCLC), pancreatic ductal adenocarcinoma (PDAC), cholangiocarcinoma (CC) and glioblastoma. Well known risk factors or carcinogenesis inducing factors are e.g. obesity, tobacco smoke, alcohol and ionizing radiation.

1.7 Transcriptional profiling in cancer cells

One of the key features in cancer cells are upregulated cell growth and prolonged viability (Hanahan and Weinberg 2011). There the key players are often mutated oncogenes coding for proteins that are involved in cell signalling leading to up- and /or dysregulated pathway signalling (e.g. PDGFR, EGFR, KRAS). These signalling pathways end in the activation of transcription factors that lead to changes in the transcription pattern of the cell. In cancer cells several transcription factor groups seem to be important. Hormone receptors as oestrogen or androgen receptors play a central role in breast or prostate cancer. STATs and c-JUN are also known transcription factors that can be persistently activated in different tumour diseases (Darnell 2002, Yeh et al. 2013). Not only persistently activated TF are important and interesting for cancer research. Changes in transcription activity patterns of cancer cells compared to normal cells during carcinogenesis (Siletz et al. 2013) or chemotherapy play a pivotal role in the understanding of cell signalling pathways. There the focus on transcription factor activity patterns and changes could also be an interesting as new therapy strategies (Darnell 2002).

1.8 Aim of the thesis

This thesis is based on the scientific results of the development of EXTs (Botvinnik et al. 2010). Here I describe further development from an oligo scalable reporter gene system to a highly scalable high throughput reporter system – called EXTassays – for the activity analysis of multiple transcription factors in parallel in living cells.

Highly complex reporter gene libraries were synthesized with an on-chip synthesis method (agilent technologies). During the cloning of the reporter gene libraries functionality experiments of the barcode system itself – the EXTs – and of the 36-mer *Cis*-element cluster were performed. In an optimisation process several different plasmid backbones were tested to improve the sensitivity and specificity.

In a proof-of-principle experiment we used *Cis*-regulatory elements to monitor the activity of transcription factors during a proliferation assay. In parallel, we measured the activity of more than thousand different transcription factor binding sites and their transcription factors and compared different cancer cell lines.

During the development of the EXTassays, several key questions were addressed:

- Is it possible to cover the complexity of the libraries during the cloning process?
- Do the *Cis*-elements show a specific activation profile in combination with the corresponding transcription factor?
- Do not specific transcription factors perform an activation via binding on the *Cis*-element?
- Is it possible to get stable and viable readouts of complex libraries of the assay in transient cell culture experiments?
- If transient cell culture experiments are not working for complex libraries, are stable cell lines the solution?

2. Materials and Methods
Materials

2.1 Laboratory material and equipment

Equipment

Arium 611 ultrapure water system	Sartorius
Axiovert 25	Zeiss
Biofuge pico	Heraeus Instruments
Biofuge fresco	Heraeus Instruments
BioPhotometer	Eppendorf
Cell Culture Hood	Heraeus Instruments
Centrifuge 5810R	Eppendorf
Concentrator 5810R	Eppendorf
Galaxy Mini	VWR
GenePulser XCell	BioRad
HeraCell 150 CO_2	Heraeus
Herasave KS 12	Heraeus
Ika Vibrax VXR	Janke&Kunkel
Ion OneTouch	Ion torrent, life technologies
Ion PGM Sequencer	Ion torrent, life technologies
Ion Proton Sequencer	Ion torrent, life technologies
Labofuge 400	Heraeus
Microplate reader Mitras LB940	Berthold Technologies
Multitron shaking incubator	Infors AG
pH meter	Sartorius
Pharmacia EPS 500/400	Pharmacia
Picodrop Spectrometer	PicodropLimited
R-202 microwave	Sharp
Sorval Ultracentrifuge	Thermo Scientific
ThermocyclerT3	Biometra
ThermocyclerT3000	Biometra
Thermomixer5436	Eppendorf
Ultra-Low Temperature Freezer U725-VIP	New Brunswick Scientific

UV-System	iNTAS
Vortex Genie2	Bender&Hobein AG
WNE 10 waterbath	memmert
XCell SureLock Mini-Cell chamber	Invitrogen

Kits

NucleoBond PC100 Kit	Macherey-Nagel (740573.100)
NucleoBond Xtra Maxi EF Kit	Macherey-Nagel (740424.10)
NucleoSpin Gel and PCR clean-up Kit	Macherey-Nagel (740609.50)
NucleoSpin Plasmid QuickPure Kit	Macherey-Nagel (740615.250)
RNAse free DNAse kit	Qiagen (79254)
RNeasy Mini Kit	Qiagen (74106)

Software

Adobe Illustrator CS3	Adobe
LaserGene 8	DNA Star Inc.
Mac OS X	Apple Inc.
Microsoft Office 2008 for Mac	Microsoft
Microwin 2000	Berthold Technologies
R (statistical computing environment)	Open Source

Plastic ware

General laboratory materials from Eppendorf, BD Falcon, Gilson, ABgene and Menzel-Gläser were used for molecular biology applications.

For cell culture applications plastic wear from BD Falcon, Eppendorf and Greiner-Nunc was used.

2.2 Reagents

General chemicals from Sigma-Aldrich or Merck were used unless stated otherwise.

Chemicals

2-Propanol	Merck
6x DNA Loading Dye	Fermentas
Agarose low EEO	AppliChem

Bacto Agar	BD
Bacto Peptone	BD
Bacto Tryptone	BD
Bacto Yeast Extract	BD
Boric acid	Merck
Bovine serum albumine (BSA)	Roche
Bromophenol blue	Merck
Chloroform	Roth
Complete tablet, Mini, EDTA-free	Roche (11 836 170 001)
Dithiothreitol (DTT)	Sigma-Aldrich
Ethanol (EtOH)	J.T.Baker
Ethidiumbromide (EtBr)	Sigma-Aldrich
Ficoll 400	Pharmacia
Flag M2 resin	Sigma-Aldrich (A2220)
Gene Ruler 50 bp DNA ladder	Fermentas
Gene Ruler 100 bp DNA ladder	Fermentas
Gene Ruler 1 kb DNA ladder	Fermentas
Glacial acetic acid	Merck
Glucose	Merck
Glycerol	Merck
Glycogen (20 mg/mL)	Roche (10 901 393 001)
Isoamylalcohol	Roth
Methanol (MetOH)	J.T.Baker
No-fat milk powder	drug store
NuPAGE 4-12%Bis-Tris Gels	Invitrogen (NP0321BOX)
Phenol	Roth
Phosphatase Inhibitor Cocktail II	Sigma-Aldrich (P5726)
PhosStop tablet	Roche
Potassiumchloride (KCl)	Merck
Potassiumhydroxide (KOH)	Merck
Power SYBR Green PCR Master Mix	Applied Biosystems
PVDF Membrane Hybond P	Amersham Biosciences
RNAse free water	Qiagen
Sodiumacetate (NaAc)	Merck

Sodiumchloride (NaCl)	Merck
Sodiumhydroxide (NaOH)	Merck
6% TBE gels	Invitrogen (EC6265BOX)
Titriplex III (EDTA)	Merck
Tris-base	Sigma-Aldrich
Tween 20	Sigma-Aldrich
Whatman paper	Sigma-Aldrich
Xylene cyanol FF	Sigma-Aldrich

Antibiotics for molecular biology (with the concentration to use)

Ampicillin	200 μg/mL	Sigma-Aldrich (A9518)
Blasticidin	75 μg/mL	Invivogen (ant-bl-1)
Chloramphenicol	50 μg/mL	Sigma-Aldrich (C0378
Zeocin	35 μg/mL	Invivogen (ant-zn-1)

Enzymes

BP clonase II	Invitrogen (11789-020)
Easy A Taq-polymerase	Stratagene (600400)
Hot StarTaq Plus	Qiagen (203643)
KOD	Novagen (71085-3)
LR clonase II	Invitrogen (11791-020)
Pfu Turbo C_x	Stratagene (600410)
Pfu Ultra High-Fidelity AD	Stratagene (600385)
Proteinase K	Invitrogen (25530-015)
Pwo Polymerase	Roche (03789403001)
Restriction Enzymes	New England Biolabs
RNase A	Invitrogen (12091-021)
RNAse free DNase	Promega (M6101)
RNAse free DNAse	Qiagen (79254)
Superscript III reverse transcriptase	Invitrogen (18080-093)
T4 DNA ligase	Promega (M1801)

Bacterial E.coli transformation competent cell strains

DH-5α (chem. comp.)	MPI
DH-5a (electro comp.)	MoBiTec (9027-TK)
DH-10b (electro comp.)	MPI
ElectroMax DH-10b (electro comp.)	Invitrogen (18290-015)
Mach1 (chem. comp.)	Invitrogen (C862003)
Top10 (chem. comp.)	Invitrogen (C4040-10)
XL-1 blue (chem. comp.)	MPI

Media and solutions for molecular biology

LB-Medium (Luria-Bertani Medium)

Per liter: dissolve in 950 mL dH_2O

YeastExtract	5 g
Bacto Peptone	10 g
Sodiumchloride (NaCl)	10 g

was adjusted to a pH to 7.0 with 5N NaOH, adjusted to a volume of 1 liter with dH_2O and then autoclaved

LB-low salt-Medium

Per liter: dissolve in 950 mL dH_2O

Yeast Extract	5 g
Bacto Peptone	10 g
Sodiumchloride (NaCl)	5 g

was adjusted to a pH to 7.0 with 5N NaOH, adjusted to a volume of 1 liter with dH_2O and then sterilized by autoclaving

SOC-Medium

Per liter: dissolve in 950 mL dH_2O

Sodiumchloride (NaCl)	0.5 g
Bacto Tryptone	20 g
Yeast Extract	5 g
Potassiumchloride (KCl) solution (250 mM)	10 mL

was adjusted to a pH to 7.0 with 5N NaOH, adjusted to a volume of 1 liter with dH_2O and then autoclaved; after autoclavating add

glucose solution (1M)	20 mL

Bacterial stock freezing medium

Glycerol	65% (v/v)
$MgSO_4$	0.1 M
Tris-HCl, pH 8.0	25 mM

the solution was autoclaved

LB-Agar plates

Yeast extract	0.5% (w/v)
Bacto Peptone pH 7.5	1% (w/v)
NaCl	1% (w/v)
Bacto Agar	1.5% (w/v)

Autoclave, cool down to 55°C in a water bath, add antibiotics and pore the plates

For blue-white selection include

X-Gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside)	35 µg/mL
IPTG (Isopropyl-β-D-thiogalactopyranosid)	15 µg/mL

dNTP mix 50x

(dATP, dCTP, dGTP, dTTP)	10 mM (2,5 mM each)
Final concentration in the PCR	200 µM (50 µM each)

Primers

Delivery concentration	50 pmol/µL
Final concentration in the PCR reaction	0.2 µM (5-10 pmol/reaction)

TAE Buffer (50x)

Tris base	242 g
Glacial acetic acid	57.1 mL
EDTA solution (0.5M pH 8.0)	100 mL

Volume was adjusted to 1 liter with dH_2O

TBE Buffer (10x)

Tris base	108 g
Boric acid	55 g
EDTA-solution (0.5M pH 8.0)	20 mL

Volume was adjusted to 1 liter with dH_2O

TE (Tris-EDTA 10x)

Tris-Base pH 7,4	10 mM
EDTA (0.5 M; pH 8)	1 mM

Volume was adjusted to 1 liter with dH_2O

Gel loading Buffer (6x)

Bromophenol blue	0.25% (w/v)
Ficoll (Type 400; Pharmacia)in H2O	15% (w/v)
Xylene cyanol FF	0.25% (w/v)

DNA extraction Buffer

Tris-HCl pH 8.0	10 mM
EDTA	100 mM
NaCl	100 mM
SDS	0.5%

Luciferase assay buffers

Firefly Luciferase Assay Buffer

Tricine	20 mM
$(MgCO_3)_4*Mg(OH)_2*5H_2O$	1.07 mM
$MgSO_4$	2.67 mM
EDTA	0.1 mM
DTT	33.3 mM

Add 0,001 V of 37% HCl to dissolve the magnesium carbonate

When the solution becomes clear adjust the pH to 7.8 using 5M NaOH

Add remaining components:

Coenzym A	270 µM

D-Luciferin, free acid	470 µM
ATP	530 µM

Renilla Luciferase Assay Buffer

NaCl	1.1 M
Na_2-EDTA	2.2 mM
$KxPO_4$ (pH 5.1)	0.22 M
BSA	0.44 mg/mL
NaN_3	1.3 mM
Adjust pH value to 5.0 and add	
Coelenterazin (dissolved in EtOH)	1.43 mM

$KxPO_4$ (pH 5.1)

Prepare 1M KH_2PO_4 solution, adjust the pH to 5.1 using 2M KOH

Western blotting buffers

RIPA Buffer

Per 250 mL dissolve in 100 mL dH_2O

Tris-HCL 1M pH 7.4	12.5 mL
NaCl 3M	12.5 mL
EDTA 0.5M, pH 8	0.5 mL
Triton-X100 (10% solution)	25 mL
sodium deoxycholate (10% solution)	25 mL
SDS (10% solution)	2.5 mL

was adjusted to a volume of 250 mL with dH_2O

Triton-X standard IP Buffer

Per 250 mL dissolve in 200 mL dH_2O

Tris 1M pH 7.5	12.5 mL
NaCl 5M	7.5 mL
Triton-X100 (10% solution)	25 mL
EGTA 0.5M, pH 8	0.5 mL

was adjusted to a volume of 250 mL with dH_2O

Triton-X cell lysis Buffer

Triton-X standard IP buffer	50 mL
Complete tablet (Roche) for 50 mL	1 piece
Zinc-chloride ($ZnCl_2$, 1M)	50 µL
Sodium-vandate (Na_3VO_4; 0.2M)	250 µL
Sodium-pyrophosphate ($Na_4P_2O_7$; 0.2M)	1125 µL
Sodium-fluoride (NaF; 0.5M)	1000 µL

Stored at 4°C

TBS (Tris-buffered Saline 20x)

Per liter: dissolve in 800 mL dH_2O

Tris-base	1 M
Sodiumchloride (NaCl)	3 M

was adjusted to a pH to 7.4 and adjusted to a volume of 1 liter with dH_2O

TBS-T (Tris-buffered saline with Tween 1x)

Per liter: dissolve in 800 mL dH_2O

TBS (20x)	50 mL
Tween20	300 µL

was adjusted to a volume of 1 liter with dH_2O

NuPAGE running Buffer MES (20x)

2-(N-morpholino)-Ethansulfonate (MES)	1 M
Tris-base	1 M
Sodiumdodecylsulfate (SDS)	2%
EDTA	20 mM

was adjusted to a pH 7.3 and a volume to 1 liter with dH_2O; store at 4°C

NuPAGE LDS sample Buffer (4x)

Tris-HCl	424 mM
Tris-base	564 mM
Lithiumdodecylsulfate (LDS)	8% (w/v)
EDTA	2.04 mM
Glycerol	40% (w/v)

ServaBlue G250 (1% solution)	7.5% (v/v)
Phenolred (1% solution)	2.5% (v/v)

Was adjusted to a volume of 10 mL and stored at -20°C

NuPAGE transfer Buffer (20x)

Bicine	500 mM
Bis-Tris (free base)	500 mM
EDTA	20 mM
Chlorobutanol	1 mM

was adjusted to a pH 7.2 and a volume to 1 liter with dH_2O; store at 4°C

NuPAGE transfer Buffer (1x)

NuPAGE transfer buffer (20x)	50 mL
Methanol	200 mL

was adjusted to a volume of 1 liter with dH_2O

Blocking Buffer

Non-fat milk powder	50 g
TBS-T (1x)	1000 mL

Stored at 4°C

2.3 Reagents for cell culture

Chemicals for cell culture

dbcAMP	BioLog, Bremen (D 009)
DMEM (Dulbeco's modified eagle medium)	Lonza (BE12-707F)
DMEM (high glucose)	Lonza (BE12-914F)
DMEM-F12 + GlutaMAX	Gibco (31331-028)
DMSO (Dimethylsulfoxide)	Sigma-Aldrich (D 8418-500ML)
Fetal Bovine Serum (FBS)	Gibco (10500-064)
GlutaMAX 100x	Gibco (35050038)
Horse Serum (HS)	Gibco (16050-122)
Hygromycin B	Gibco (10687-010)
Leibovitz's L-15 Medium	Gibco (11415-049)

Lipofectamine 2000	Invitrogen (11668-019)
McCoy`s 5A Medium + GlutaMAX	Gibco (36600-088)
MEM (Minimum Essential Media)+ GlutaMAX	Gibco (42360-024)
Opti-MEM	Gibco (31985-047)
Penicillin/Streptomycin (Pen/Step)	Lonza (17-602E)
PMA (phorbol 12-myristate 13-acetate)	Sigma-Aldrich (P8139)
Polyethylenimine (PEI)	Sigma-Aldrich (408727)
Poly-L-Lysine (PLL)	Sigma-Aldrich (P4707)
Puromycin	life technologies (A1113802)
RPMI-1640 Medium + GlutaMAX	Gibco (61870-010)
Trypsine 10x	Lonza(BE17-160E)

Media and solutions for cell culture

Freezing medium for eukaryotic cell lines

DMEM	60%
DMSO	5-10%
Fetal bovine serum (FBS)	30%

293HEK growth-medium (also for 293HEK_FT)

DMEM (high Glucose)	450 mL
Fetal Bovine Serum 10% f.c.	50 mL
Pen/Strep 1% f.c.	5 mL
GlutaMAX (L-Glutamine) 1% f.c.	5 mL

Caco-2 growth medium

MEM (Minimum Essential Media)+GlutaMAX	400 mL
Fetal Bovine Serum 20% f.c.	100 mL
Pen/Strep 1% f.c.	5 mL

LS411N growth medium (also for LS513, LS1034)

RPMI-1640 Medium + GlutaMAX	450 mL
Fetal Bovine Serum 10% f.c.	50 mL
Pen/Strep 1% f.c.	5 mL

MCF-7 growth medium (also for HeLa, Hs 633T, HT-1080)

DMEM (low Glucose)	450 mL
Fetal Bovine Serum 10% f.c.	50 mL
Pen/Strep 1% f.c.	5 mL
GlutaMAX (L-Glutamine) 1% f.c.	5 mL

PC12 growth-medium

DMEM (low Glucose)	450 mL
Horse Serum 5% f.c.	25 mL
Fetal Bovine Serum 10% f.c.	50 mL
Pen/Strep 1% f.c.	5 mL
GlutaMAX (L-Glutamine) 1% f.c.	5 mL

SH-SY5Y growth medium

DMEM-F12 + GlutaMAX	450 mL
Fetal Bovine Serum 10% f.c.	50 mL
Pen/Strep 1% f.c.	5 mL

SW403 growth medium (also for SW480, SW620, SW837, SW1116, SW1463)

Leibovitz's L-15 Medium	450 mL
Fetal Bovine Serum 10% f.c.	50 mL
Pen/Strep 1% f.c.	5 mL

TE-671 growth medium

DMEM (low Glucose)	400-425 mL
Fetal Bovine Serum 15-20% f.c.	75-100 mL
Pen/Strep 1% f.c.	5 mL
GlutaMAX (L-Glutamine) 1% f.c.	5 mL

U-2OS growth medium (also for A-204, HT-29)

McCoy's 5A + GlutaMAX	450 mL
Fetal Bovine Serum 10% f.c.	50 mL
Pen/Strep 1% f.c.	5 mL

WiDr growth medium

MEM (Minimum Essential Media)+GlutaMAX	450 mL
Fetal Bovine Serum 10% f.c.	50 mL
Pen/Strep 1% f.c.	5 mL

Phosphate-buffered Saline (PBS) 10x

Sodiumchloride (NaCl)	100 g
Potassiumchloride (KCl)	2.5 g
Disodiumhydrogenphosphate (Na_2HPO_4 x $2H_2O$)	7.2 g
Potassiumdihydrogenphosphate (KH_2PO_4)	2.5 g

Disolve, adjust pH to 7.2 with NaOH and add H_2O to 1000 mL and then sterilized by autoclaving

Mammalian cell lines

293HEK	Human embryonal kidney (ATCC) (Graham et al. 1977)
293HEK_FT	Human embryonal kidney cells expressing SV40 large T-antigen (Invitrogen)
A-204	Human Rhabdomyosarcoma (ATCC) (Giard et al. 1973)
Caco-2	Human colorectal adenocarcinoma (ATCC) (Fogh et al. 1977)
HeLa	Human cervix adenocarcinoma (ATCC) (Scherer 1954)
Hs 633T	Human fibrosarcoma (ECACC)
HT-29	Human colorectal adenocarcinoma (ATCC) (Fogh et al. 1977)
HT-1080	Human fibrosarcoma (ATCC) (Rasheed et al. 1974)
MCF-7	Human adenocarcinoma of the mammary gland metastasis (ATCC) (Soule et al. 1973)
LS411N	Human colorectal carcinoma of the cecum, stage Duke's B (ATCC) (Suardet et al. 1992)
LS513	Human colorectal carcinoma of the cecum, stage Duke's C (ATCC) (Suardet et al. 1992)
LS1034	Human colorectal carcinoma of the cecum, stage Duke's C (ATCC) (Suardet et al. 1992)
PC12 tet OFF	rat pheochromocytoma cell line stably expressing tetracycline-controlled transactivator (tTA) under neomycine resistance. (Clontech)(Greene and Tischler 1976) (ATCC)

SH-SY5Y	Human neuroblastoma metastasis (ATCC) (Biedler et al. 1978)
SW403	Human colon adenocarcinoma, stage Duke's C (ATCC) (Leibovitz et al. 1976)
SW480	Human colon adenocarcinoma, stage Duke's B (ATCC) (Leibovitz et al. 1976)
SW620	Human colon adenocarcinoma metastasis, stage Duke's C (ATCC) (Leibovitz et al. 1976)
SW837	Human rectal adenocarcinoma, stage IV (ATCC) (Leibovitz et al. 1976)
SW1116	Human colon adenocarcinoma, stage Duke's A (ATCC) (Leibovitz et al. 1976)
SW1463	Human rectum adenocarcinoma, stage Duke's C (ATCC) (Leibovitz et al. 1976)
TE-671	Human Rhabdomyosarcoma (CLS) (McAllister et al. 1977)
U-2 OS	Human Osteosarcoma (ATCC) (Pontén and Saksela 1967)
WiDr	Human colon adenocarcinoma (ATCC) (Noguchi et al. 1979)

2.4 Vectors and Plasmids

construct	antibiotic resistance
pDONR/ Zeo	Zeocin, Chloramphenicol(Cm)
pDEST_GL3	Ampicillin, Chloramphenicol
pDEST_GL4.14_ß-globin	Ampicillin, Chloramphenicol
pDEST_Lenti-promoter	Ampicillin, Cm, Blasticidin
pENTR_TF library	Zeocin
pENTR_Xie library	Zeocin
PENTR_Co library	Zeocin
pEXPR_GL3_TF library	Ampicillin
pEXPR_GL3_Xie library	Ampicillin
pEXPR_GL3_Co library	Ampicillin
pEXPR_GL4.14_ß-globin_TF library	Ampicillin
pEXPR_GL4.14_ß-globin_Xie library	Ampicillin
pEXPR_GL4.14_ß-globin_Co library	Ampicillin
pEXPR_Lenti-promoter_TF library	Ampicillin, Blasticidin
pEXPR_Lenti-promoter_Xie library	Ampicillin, Blasticidin

pEXPR_Lenti-promoter_Co library	Ampicillin, Blasticidin
pEXPR_TripZ-promoter_TF library	Ampicillin, Zeocin, Puromycin
pEXPR_TripZ-promoter_Xie library	Ampicillin, Zeocin, Puromycin
pEXPR_TripZ-promoter_Co library	Ampicillin, Zeocin, Puromycin
pGEM-T (Promega)	Ampicillin
pGEM-T_ß-globin	Ampicillin

2.5 Primers

Cloning primers

16128	TF_F	CTTGGACAGGGTGGTGGG
16129	X_F	CGAGCGCTTCCGGTAAGA
16130	Co_F	AGAGCCACGGGCGAAAAG
117804	TF_R	ACTGCTGTCCGCTCTGCC
117805	X_R	CCACACCGACATGGGGAG
117806	Co_R	TACCTGGAATGGGGCAGC
16137	TF_Dec	CCTCCCCGATGAATTGCA
16138	X_Dec	GACACAGTGGCGCAGTGG
16139	Co_Dec	ACGCTCACCCCGAGAATG
16501	B1_TF_F	GGGGCAAGTTTGTACAAAAAAGCAGCTTGGACAGGGTGGTGGG
16502	B1_X_F	GGGGCAAGTTTGTACAAAAAAGCAGCGAGCGCTTCCGGTAAGA
16503	B1_Co_F	GGGGCAAGTTTGTACAAAAAAGCAGAGAGCCACGGGCGAAAAG
117912	B2_TF_R	GGGGCCACTTTGTACAAGAAAGCTGACTGCTGTCCGCTGTGCC
117913	B2_X_R	GGGGCCACTTTGTACAAGAAAGCTGCCACACCGACATGGGGAG
117914	B2_Co_R	GGGGCCACTTTGTACAAGAAAGCTGTACCTGGAATGGGGCAGC
16495	T3_TF_F	AATTAACCCTCACTAAAGGGCTTGGACAGGGTGGTGGG
16496	T3_X_F	AATTAACCCTCACTAAAGGGCGAGCGCTTCCGGTAAGA
16497	T3_Co_F	AATTAACCCTCACTAAAGGGAGACCCACGGGCGAAAAG
117909	T7_TF_R	TAATACGACTCACTATAGGGACTGCTGTCCGCTCTGCC
117910	T7_X_R	TAATACGACTCACTATAGGGCCACACCGACATGGGGAG
117911	T7_Co_F	TAATACGACTCACTATAGGGTACCTGGAATGGGGCAGC

17749	ß-globin intron	ATATAAGCTTGAGAACTTCAGGGTGAGTTTGGGG
17750	ß-globin rev.	ATATAAGCTTGTTGCCCAGGAGCTGTAGGAAAAAG
18225	pDEST_GL4 rev	CGCAAACGGATCCTTATCGATTTTACCAC
18226	pDEST_GL4	ATATACGCGTCGAGGATATCAACAAGTTTGTACAAAAAAGCTG
21015	pLenti ClaI rev	ATAATTAACGCGTAAGCTTATCGATACCGTCGAGA
21016	pLenti MluI rev	CGCGGTTCGAAGGTAAGCCTATCC
23984	ccdB XbaI for	AATTTACCGTTCTAGAATCAACAAGTTTGTA
23985	ccdB MluI rev	TAATATAATACGCGTATCAACCACTTTGTA
24125	ccdB rev	AACCACTTTGTACAAGAAAGCTGAAC
24277	ccdB NheI for	AATTTAGCTAGCCCCGGCCGCCATGGCCGCGG
25170	ccdB ClaI for	AATTAATCGATTCCCGGCCGCCATGGCCGCGG
24279	ccdB reverse	CGGCCGCACTAGTGATTTCTAGAATCAACCACTTTGTAC

25124	B1_CMV for	GGGGACAAGTTTGTACAAAAAAGCAGGCTCTCCGCCATGCATTAGTTATTAATAG
25125	B2_EYFP_MCS rev	GGGGACCACTTTGTACAAGAAAGCTGGGTCGATCAGTTATCTAGATCCGGTG

29391	B1 Sharp1 aa1 for	GGGGACAAGTTTGTACAAAAAAGCAGGCTCCGACGAAGGAATCCCTC
29392	B2 Sharp1 aa98 rev	GGGGACCACTTTGTACAAGAAAGCTGGGTCTCACGCTTTCAAGTGCTT
29393	B2 Sharp1 aa185 rev	GGGGACCACTTTGTACAAGAAAGCTGGGTCTCACCCAGAGCCCC
29394	B1 S1-bHLH aa88 for	GGGGACAAGTTTGTACAAAAAAGCAGGCTCCGTCTTGGAATTAACTTTAAAGCA
29395	B2 Sharp1 aa410 rev	GGGGACCACTTTGTACAAGAAAGCTGGGTCTCAGGGGGCGTCCTT
29396	B1 Sharp2 aa1 for	GGGGACAAGTTTGTACAAAAAAGCAGGCTCCGAACGGATCCCCAGC
29397	B2 Sharp2 aa122 rev	GGGGACCACTTTGTACAAGAAAGCTGGGTCTCACTGCAGGGCAATGAT
29398	B2 Sharp2 aa195 rev	GGGGACCACTTTGTACAAGAAAGCTGGGTCTCAAGCCGAGTCCAATG
29399	B1 S2-bHLH aa113 for	GGGGACAAGTTTGTACAAAAAAGCAGGCTCCCAGCAGCAGCAGAAAAC
29400	B2 Sharp2 aa411 rev	GGGGACCACTTTGTACAAGAAAGCTGGGTCTCAGTCTTTGGTTTCTAAGTTT

Primers for Next Generation Sequencing

18236	pGL4 RNA rev	ATGGTGGCTTTACCAACAGTACCG
18237	pGL4 DNA rev	AACAATCAAGGGTCCCCAAACTCA
18238	TF TATA Dec for	GCTCCTCCCCGATGAATTGC
18239	Xie TATA Dec for	CAGCTGACACAGTGGCGCAGT
18240	Co TATA Dec for	CTAGCTACGCTCACTCCGAGAATG
27505	TF Dec rev	ACTGCTGTCCGCTCTGCC

27506 Xie Dec rev GCCACACCGACATGGGGAG

27507 Co Dec rev GCTGTACCTGGAATGGGGCAGC

27740 PGM_A_IXcode001_Xie_s
CCATCTCATCCCTGCGTGTCTCCGACTCAGCTAAGGTAACGATCAGCTGACACAGTGGCGCAG

27741 PGM_A_IXcode002_Xie_s
CCATCTCATCCCTGCGTGTCTCCGACTCAGTAAGGAGAACGATCAGCTGACACAGTGGCGCAG

27742 PGM_A_IXcode003_Xie_s
CCATCTCATCCCTGCGTGTCTCCGACTCAGAAGAGGATTCGATCAGCTGACACAGTGGCGCAG

27743 PGM_A_IXcode004_Xie_s
CCATCTCATCCCTGCGTGTCTCCGACTCAGTACCAAGATCGATCAGCTGACACAGTGGCGCAG

27744 PGM_A_IXcode005_Xie_s
CCATCTCATCCCTGCGTGTCTCCGACTCAGCAGAAGGAACGATCAGCTGACACAGTGGCGCAG

27745 PGM_A_IXcode006_Xie_s
CCATCTCATCCCTGCGTGTCTCCGACTCAGCTGCAAGTTCGATCAGCTGACACAGTGGCGCAG

27746 PGM_A_IXcode007_Xie_s
CCATCTCATCCCTGCGTGTCTCCGACTCAGTTCGTGATTCGATCAGCTGACACAGTCGCGCAG

27747 PGM_A_IXcode008_Xie_s
CCATCTCATCCCTGCGTGTCTCCGACTCAGTTCCGATAACGATCAGCTGACACAGTGGCGCAG

27748 PGM_trP1_Xie_as CCTCTCTATGGGCAGTCGGTGAGCCACACCGACATGGGGAG

27749 PGM_A_IXcode001_TF_s
CCATCTCATCCCTGCGTGTCTCCGACTCAGCTAAGGTAACGATGCTCCTCCCCGATGAATTGC

27750 PGM_A_IXcode002_TF_s
CCATCTCATCCCTGCGTGTCTCCGACTCAGTAAGGAGAACGATGCTCCTCCCCGATGAATTGC

27751 PGM_A_IXcode003_TF_s
CCATCTCATCCCTGCGTGTCTCCGACTCAGAAGAGGATTCGATGCTCCTCCCCGATGAATTGC

27752 PGM_A_IXcode004_TF_s
CCATCTCATCCCTGCGTGTCTCCGACTCAGTACCAAGATCGATGCTCCTCCCCGATGAATTGC

27753 PGM_A_IXcode005_TF_s
CCATCTCATCCCTGCGTCTCTCCGACTCAGCAGAAGGAACGATGCTCCTCCCCGATGAATTGC

27754 PGM_A_IXcode006_TF_s
CCATCTCATCCCTGCGTGTCTCCGACTCAGCTGCAAGTTCGATGCTCCTCCCCGATGAATTGC

27755 PGM_A_IXcode007_TF_s
CCATCTCATCCCTGCGTGTCTCCGACTCAGTTCGTGATTCGATGCTCCTCCCCGATGAATTGC

27756 PGM_A_IXcode008_TF_s
CCATCTCATCCCTGCGTGTCTCCGACTCAGTTCCGATAACGATGCTCCTCCCCGATGAATTGC

27757 PGM_trP1_TF_as CCTCTCTATGGGCAGTCGGTGACTGCTGTCCGCTCTGCC

Sequencing primers

10764	LUCI_seq	GGCGTCTTCCATGGTGGCTTTACC
17303	pGL3 primer	GGCTGTCCCCAGTGCAAGTGCA
17358	pDONR-Zeo	CGGCAACTGCGTGCACTTCGT
18305	GL4_bGlob for	CTAGCAAAATAGGCTGTCCCCAGT
18306	GL4_bGlob rev	ACACCCTGAAAACTTTGCCCCCTC
18537	GL4_bGlob for	TTCGTCTCAGCCAATCCCTGGG
18538	GL4_bGlob for	CGCATGATGACCACCGATATGGC
18539	GL4_bGlob for	CTACACCCTGGTCATCATCCTGC
18540	GL4_bGlob for	GAACAGTAGTGGCAGTACCGGATTG
18541	GL4_bGlob for	AACACCCCAACATCTTCGACGCC
18542	GL4_bGlob for	AGTCAGCAACCATAGTCCCGCC
18543	GL4_bGlob for	CATCACCGCCGTAATCGACTGG
18544	GL4_bGlob for	GGCGGTAATACGGTTATCCACAGAA
18545	GL4_bGlob for	TGATCAGTGAGGCACCGATCTCAG
21792	pLenti for	CGCCTCCATCCAGTCTATTAATTGTTGC

Methods

2.6 Molecular biology methods

Polymerase Chain Reaction (PCR)

The Polymerase Chain Reaction is a common method for in vitro amplification of a defined DNA sequence. The PCR was developed by Kary Banks Mullis (Mullis 1990; Mullis and Faloona 1987). For the amplification of the DNA a thermostable DNA polymerse is needed. This DNA polymerase (Taq polymerase) derived from the thermophilic bacterium Thermus aquaticus amplifies selected template DNA molecules multiple times. In a PCR reaction two DNA oligonucleotides (primers) that bind complementary to flank the strech of the DNA template are needed because the Taq polymerase is unable of de novo DNA synthesis. After primer annealing the templates can be extended by the enzyme to make a full-length PCR product. Both strands of the DNA are replicated in every cycle of amplification. One of the primers always binds as a sense primer on the plus strand and the other one as an anti-sense primer to the minus strand. The newly generated PCR products from every round of amplification serve as DNA templates leading to exponential amplification kinetics. Taq polymerases make about 0.8 mismatches per 1 kb per amplification cycle. To improve the DNA quality for the amplification of long fragments we used DNA polymerases with a 3'-5' exonuclease-dependent proofreading function. DNA polymerases with this nature are Pfu polymerases (Stratagene), derived from the thermophilic archaea bacterium Pyrococcus furiosus or Pwo polymerases (Roche), derived from the thermophilic archaea bacterium Pyrococcus woesei.

Composition of the PCR reaction:	final concentration
10x reaction buffer	
template DNA	500 pg-5 ng/reaction
$MgCl_2$ (included in the buffer)	1-4 mM
Deoxynucleotide (dATP, dCTP, dTTP, DGTP)	0.2 mM each
Forward primer (5-10 pmol/reaction)	0.2 µM
Reverse primer (5-10 pmol/reaction)	0.2 µM
Taq polymerase (0.5-1 U/reaction)	0.01 U/µL
H_2O	

A PCR reaction is composed of three main steps:

1. Melting of the DNA

2 .Primer annealing and

3. Elongation.

The PCR steps run at different temperatures and are repeated in cycles multiple times. For desinging PCR programmes the following rules were regarded:

Initial denaturation: 2-5 min at 95°C was used to remove secondary DNA structures. This step was extended up to 15 min when a hot-start Taq polymerase (Qiagen) was used. Hot-start polymerases need prolonged initial denaturation steps to get aktivated.

Denaturation: 30 sec at 95°C; In this first step of a cycle the DNA strands are melted and become available for the primers.

Annealing: 20-30 sec at a temperature 4 degrees lower than the calculated primer Tm. In some cases a gradient PCR machine was used for optimization of the annealing temperature.

Elongation: at 72°C – the optimal activity temperature for Taq polymerases. The elongation time was calculated depending on the size of the PCR product and the enzyme used. Taq polymerases need roughly 30 sec elongation time per kb and DNA polymerases with proofreading activity need 1 min/kb.

Final extention: when the whished number of cycles is completed, the PCR will further be incubated 5-10 min at 72°C to ensure that smaller PCR products that could be produced due to incomplete elongation will be filled up.

Designing primer sequences

Primer sequences were designed to have a melting temperature between 54° and 62°C. In some cases lower and higher temperatures had to be used. With the following formula the melting temperature of the primer can be roughly calculated:

$Tm = (A+T)x2 + (G+C)x4$,

In the formula A, T, C and G stand for the number of the corresponding nucleotide in the primer sequence. For a more precise Tm calculation we used an online algorithm that was developed by Warren A. Kibbe (Kibbe 2007), online available at http://www.basic.northwestern.edu/biotools/oligocalc.html

Cloning of PCR products via pGEM-T cloning

The pGEM-T vector from Promega is a linearized cloning-vetor with a single 3'–terminal thymidine overhang. PCR products that were amplified with Taq polymerase (DNA-Ploymerse from bacteria strain Thermus aquaticus), were cloned into the pGEM-T vector. The procedure is based on the nature of the Taq polymerase often to add template-independent a single deoxyadenosine to the 3'–ends of the PCR product. This aspect leads to a complementary binding between PCR-product and cloning vector. Therefore an efficient ligation, catalyzed by the T4 DNA ligase (Promega) is enabled.

Classical cloning

When the gene of interest is cloned into the pGEM-T transfer vector, the gene of interest was cut out for further cloning with the help of restriction enzymes: Type II DNA endonucleases from New England Biolabs. The vector of interest was digested with the same enzymes or enzymes that produce compatible ends as the gene of interest. After the incubation the DNA fragments were separated by gel-electrophoresis. The agarose gels contained 0.01% ethidium bromide (EtBr) in 1x TAE buffer. The DNA fragments were visualized via the intercalating nature of EtBr with UV light. In order to extract the elected DNA fragments, the bands with the correct length of base pairs were excised from the gel and purified with the help of the NucleoSpin Gel and PCR Clean-up kit (Macherey-Nagel). Cleaned up fragments (insert and vector backbone in a molar ratio 3:1) were mixed with 1 µL of ligation buffer in 10x concentration and ligated with the help of 1 µL T4 DNA ligase (Promega) in a final reaction volume of 10 µL. After sufficient incubation the ligation mix was transformed into E. coli strain XL-1 blue, DH-5α, DH-10b (self-made competent bacteria) or commercial available E. coli strains like Mach1, MegaX, TOP10, Stbl3 (all from Invitrogen, life technologies).

DNA digest with restriction endonucleases

For DNA restriction digests type II restriction endonucleases were used. Type II restriction endonucleases induce cleavage mostly within their palindromic recognition sites of 4-8 nucleotides. This cleavage result in 5'- or 3'-DNA overhangs (sticky ends) or blunt ends depending on the enzyme. In a digestion reaction with restriction enzymes 2-5 µg plasmid DNA was applied to 2-5 units of the enzyme in a reaction volume of 40-60 µL. The restriction enzyme activity unit 'U' defines the amount of enzyme that is needed to digest 1 µg DNA in 1 hour. Compatible reaction buffer in 10x concentration was selected according to the supplier's recommendations. In most cases the incubation time was 1 hour at 37°C.

Dephosphorylation of 5'-DNA fragment overhangs, vectors only

After a DNA restriction digest of vector plasmid the DNA was treated with calf intestinal alkaline phosphatase (CIP, Roche). The enzyme CIP removes the terminal 5'-phosphate groups of dsDNA. Through this treatment the unwanted re-ligation of vectors is reduced. To a completed DNA digest (e.g. volume 40 µL) 6 µL CIP 10x buffer, 12 µL H_2O and 2 µL enzyme was added. This mix was incubated for 20 min at 37°C. To stop the reaction the DNA was separated by agarose electrophoresis.

Agarose gel electrophoresis

For the separation of DNA fragments between 0.1 kb and 10 kb agarose gels of 1% - 5% were used. To 1x TAE buffer the proper amount of agarose was added. The mix of buffer and agarose was boiled in the microwave at 800 W until the suspension was clear; this took approximately 5-10 min. Then the liquid agarose solution was cooled for 15 min to avoid toxic fumes of EtBr. When adding the EtBr into the gel at too high temperatures the EtBr evaporates as toxic bromine gas. After cooling down the agarose gel EtBr was added to a final concentration of 1µg/µL. into a gel-casting form with combs that define the pockets for sample loading the fluid agarose was poured. For solidification the gel needed 20-60 min depending on the concentration of agarose and room temperature. Into a chamber with buffer 1x TAE the gel was placed. The DNA samples were prepared with 10x loading dye and loaded into the pockets of the gel. Depending on the size of the electrophoresis chamber and the agarose concentration the voltage applied was 140-180 V.

The migration in the agarose gel of DNA fragments is towards the positive electrode because the phosphates in the DNA backbones are negatively charged.

Depending on the length of base pairs of the DNA fragments and the obstruction by the fishnet like structure of the agarose gel the speed of the migration is lowered. This results in the size-dependent separation of DNA fragments by gel electrophoresis. The DNA fragments were visualized by the fluorescence of the intercalated EtBr when exposed to a UV light source emitting 260 nm.

The O'GeneRuler 50bp DNA ladder, O'GeneRuler 100bp DNA ladder and O'GeneRuler 1kb (Fermentas, Thermo Scientific) were used as molecular size standards. Gel chamber, combs and casting form were produced by the Max-Planck-Institute fine mechanics facility.

Isolation of DNA from agarose gels

After identification of the desired DNA band in the agarose gel, the DNA fragment was excised under UV light (356 nm, Intas UV systems) by scalpel in order to purify the selected DNA. To avoid DNA crosslinks caused of UV this excision should be as fast as possible. The gel particle was transferred into an Eppendorf reaction tube and weighted. Then the DNA was purified with the NucleoSpin Gel and PCR Clean-up kit (Macherey-Nagel) according to the manufacturer's protocol. To solve the agarose to each 100 mg gel 200 µL buffer NTI was added and incubated for 5-10 min at 50°C with moderate shaking. When the gel was dissolved completely the DNA was loaded onto a silica membrane, washed two times with 600 µL buffer NT3 and dried by centrifugation. Depending on the amount of DNA seen on the gel previously the DNA was eluted in 30-50 µL buffer NE. After elution the size and quality of the purified DNA fragment was checked via gel electrophoresis and the concentration was determined via Spectrophotometer.

Ligation of DNA fragments

DNA molecules were ligated by using T4 DNA ligase (Promega). This enzyme derived from the bacteriophage T4 catalyzes the formation of covalent phosphodiester bonds between free 3'-hydroxy and 5'-phosphate overhangs of double stranded DNA. ATP and Mg^{2+} are cofactors needed for this reaction. The cloning reaction of DNA fragments, that is the ligation of an insert and a vector, incubated for 1 hour at room temperature or at 4°C overnight.

Transformation of chemically competent bacteria

An aliquot of chemically transformation competent E. coli, strain XL1 blue, DH5α, or Mach1 was thawed o ice. 2-20 µL ligation mix or recombination reaction was added to 20-100 µL bacteria. This transformation mix was incubated for 20 min on ice. After incubation the bacteria were heat-shocked at 42°C for 44 sec and immediately put back on ice for 2 min. Then 800 µL of SOC-Medium was added and the bacteria incubated at 37°C for 1 hour with shaking (160 rpm). When plasmids were transformed that contained retroviruses (pLenti or pTripZ) or adeno-associated viruses (AAV) the incubation steps after transformation were done at 30°C. During the incubation step the bacteria start expressing the appropriate resistance against antibiotic. After incubation the bacteria were centrifuged and resolved in 50 µL LB-medium. This volume was plated on pre-warmed LB-agar plates containing the appropriate antibiotics. For equally distribution of the bacteria sterile glass bullets were used. The plates were incubated over night at 30-37°C, depending on the encoded genes (viral genes should be expressed on 30°C for two days).

Transformation via electroporation of bacteria

For the electroporation the bacteria (E. coli strain DH-10b, MegaX (Invitrogen)) were thawed and diluted 1:2 with sterile 10% glycerol. Then 2-2.5 µL of ligation mix or recombination reaction was added to the bacteria. The transformation mix was transferred into 1mm electroporation cuvettes (BioRad).

With the following settings the electroporation was done with 'GenePulserII'(BioRad): 1.75 kV, 25 µF capacitance and 200 Ω resistance. The bacteria were resuspended in 800 µL SOC-medium and incubated at 30-37°C for 1 hour with shaking (160rpm). After incubation the bacteria were treated as described above in section 'Transformation of chemically competent

bacteria'.

Plasmid DNA purification

The two main steps at plasmid DNA preparations are the alkaline cell lysis followed by SDS precipitation of genomic DNA and proteins. The plasmid DNA preparation is based on a modified protocol of Birnboim and Doly (Birnboim and Doly 1979). The plasmid DNA bound on a silica column under appropriate high-salt conditions, washed and eluted in buffer TE. Plasmid DNA purification was extracted from fresh bacteria cultures, grown overnight in LB-medium with appropriate antibiotics. For the preparation plasmid DNA purification Kits with different scales (Macherey-Nagel) were used. Detailed description of the preparation is available in the manufacturer's protocol.

Plasmid DNA mini preparation

For plasmid DNA purification in small scale (4mL) the NucleoSpin Plasmid QuickPure kit (Macherey-Nagel) was used. In brief, 2-4 mL bacterial culture was pelleted and resuspended in 250 µL buffer A1 with RNAse A. Cell lysis was done with 250 µL buffer A2, this reaction was stopped with 300 µL pre-cooled buffer A3. To pellet the genomic DNA and proteins the mix was centrifuged for 10 min. The supernatant was loaded to the silica column, washed with 600 µL buffer AQ, dried by centrifugation and eluted in 100 µL buffer TE.

Plasmid DNA midi preparation

For plasmid DNA purification in middle scale (100 mL) the NucleoBond PC 100 kit (Macherey-Nagel) was used. In brief, 100 mL bacterial culture was pelleted and resuspended in 4 mL buffer S1 with RNAse A. Cell lysis was done with 4 mL buffer S2, this reaction was stopped with 4 mL pre-cooled buffer S3. The silica columns were equilibrated with 2.5 mL buffer N2. The lysate was clarified through pressing into a folded filter. The cleaned lysate was loaded to the silica column, washed with 10 mL buffer N3, dried and eluted in 5 mL buffer N5. The eluted plasmid DNA was precipitated with 3.5 mL isopropanol. The DNA pellet was resolved in 100 µL buffer TE. After DNA concentration determination the DNA concentration was adjusted to 1 µg/µL.

Plasmid DNA maxi preparation

For endotoxin free plasmid DNA purification in big scale (200-300 mL) the NucleoBond Xtra Maxi EF kit (Macherey-Nagel) was used. In brief, 200-300 mL bacterial culture was pelleted and resuspended in 12 mL buffer RES-EF with RNAse A. Cell lysis was done with 12 mL buffer S2, this reaction was stopped with 12 mL pre-cooled buffer S3. The silica columns were equilibrated with 2.5 mL buffer N2. The lysate was clarified through pressing into a folded filter. The cleaned lysate was loaded to the silica column, washed two times with 10 mL buffer N3 and two times with 8 mL buffer N4, dried and eluted in 5 mL buffer N5. The eluted plasmid DNA was precipitated with 3.5 mL isopropanol. The DNA pellet was resolved in 100 µL endotoxin free buffer TE. After DNA concentration determination the DNA concentration was adjusted to 1 µg/µL.

Generating frozen stocks of bacterial cultures

To avoid the time consuming process of re-transformations of plasmids into bacteria, frozen stocks were prepared. After cloning and sequence verification of constructs, 2 mL of fresh bacterial culture were pelleted by centrifugation. The pellet was resuspended in 1 mL of bacterial freezing medium. The samples were shock frosted in liquid nitrogen and stored at -80°C. Viable bacteria were recovered from the frozen stocks by scratching the frozen surface of the sample with a pipette tip and inoculating LB cultures.

Photometric concentration determination of nuclear acids

The purity and concentration of nucleic acid solution can be analyzed by spectrophotometry. The concentration of a solution is directly proportional to its extinction or absorption, according to the Lambert-Beer law, that is

$A = \varepsilon * c * l$

With ε, that is the molar extinction coefficient (unit M-1 cm-1), c representing the concentration (unit M) and l being the optic path length (cuvette thickness) where the light passes through the sample (unit cm). The extinction coefficients of nuclear acids at $\lambda = 260$ nm are:

Guanine: $\varepsilon = 12010$ M-1 cm-1

Cytosine: $\varepsilon = 7050$ M-1 cm-1

Adenine: $\varepsilon = 15200$ M-1 cm-1

Thymine: $\varepsilon = 8400$ M-1 cm-1

(Uracile: $\varepsilon = 8111$ M-1 cm-1)

The absorption at 260 nm (maximum of absorption of nucleic acids) and 280 nm (maximum of absorption of aromatic amino acids in proteins) of the nucleic acid solution was determined by using a spectrophotometer. For a reliable measurement it is really important that the value for the absorption at 260 nm is arranged between 0.1 and 1 to fit the linear range of the spectrophotometer. Therefore a dilution of the sample was necessary. For the determination of the concentration the following relation was applied:

1 OD260 (optical density at 260 nm) = 50 mg/mL for double stranded (ds) DNA

1 OD260 = 40 mg/mL for single stranded (ss) RNA

1 OD260 = 33 mg/mL for ss oligonucleotides

For the determination of the concentration of ds DNA following equation was used:

C (µg/µL) = OD260 * 50 * dilution factor/1000.

The purity of a sample was assessed by the relative absorption values at 260 nm over 280 nm. Values between 1.8 and 2.0 exist in a clean DNA or RNA preparation. Significant lower values indicated a protein contamination of the sample.

DNA sequencing

DNA sequencing was done at the Institute's DNA Core Facility (Department of Neurobiology, MPI of Experimental Medicine, Göttingen) lead by Fritz Benseler. DNA sequencing is based on a modified dideoxy chain reaction termination method according to Sanger (Sanger et al. 1977). The procedure is based on a linear DNA amplification in the presence of DNA polymerase, a sequencing primer, deoxynucleoside triphosphates (dNTPs), and modified nucleotides (dideoxyNTPs), that terminate the DNA strand elongation. DideoxyNTPs stochastically embedded into the newly synthesized DNA terminating the synthesis by replacing the normal dNTPs. The PCR products differ in their length and were separated by capillary electrophoresis. As each of the four dideoxyNTPs is coupled to a different fluorescent dye the DNA fragments can be characterized by size, and the terminal nucleotide can be identified. The DNA sequence is determined through a repetitive procedure.

Oligonucleotide synthesis

Oligonucleotide synthesis was done at the DNA Core Facility (Max-Planck-Institute for Experimental Medicine).

Analysis of DNA sequences

Using the DNAstar software, version 8, provided by Lasergene the sequencing data were analysed. The sequences were also aligned to databases of the 'National Center for Biotechnology Information' (NCBI, http://www.ncbi.nlm.nih.gov).

DNA precipitation

For concentration of DNA samples, DNA solutions were precipitated. The volumes of the samples were adjusted to 100 µL. Then 1 µL of glycogen or pellet paint was added for visualizing the precipitated DNA. After one vortexing step 10 µL of 3M Sodium-Acetate pH 5.2 (NaAc) was added and the sample was mixed via vortexing. Then 250 µL of 100% ethanol was added and mixed via vortexing again. The sample was incubated at -20°C for 1 hour and pelleted by centrifugation at 13.000 rpm for 20 min. The pellet was washed two times with 80% ethanol and centrifugation at 13.000 rpm for 5 min each. After a drying step the sample was resolved in a desired volume of buffer TE (pH 7.5).

Gateway recombination cloning – One way Gateway cloning

Classical cloning relies on type II DNA endonucleases and T4 DNA ligases divided in several steps with suboptimal efficiancy. Due to this fact it is a time consuming process. In contrast to classical cloning the Gateway cloning technology from Invitrogen (Invitrogen, life technology) is based on site-directed recombination enzymes derived from the bacteriophage lambda, that are able to transfer DNA fragments between vectors (Hartley et al. 2000), and more efficient. Detailed information about the Gateway cloning technology can be found online in the manufacterer's protocol. The desired DNA sequence was ampilfied via PCR in order to clone these product. The DNA primer carry two specific recombination sites: attB1 and attB2. In a so called 'BP raction', catalysed by the Gateway BP Clonase II enzyme mix, containing the viral recombination enzyme Integrase, the E. coli-encoded protein Integration

Host Factor and the reaction buffer, the PCR product was recombined with the pDONR vector carrying attP1 and attP2 recombination sites. Each BP reaction included:

attB-PCR product	3 µL (PCR purified, 100 ng)
pDONR_Zeo	1 µL (100 ng)
BP clonase II	1 µL
Total reaction volume	5 µL

The incubation of the reaction was at room temperature for 1 hour or at 4°C overnight. After the sufficient incubation 1 µL was used for a electroporation of DH-10b (selfmade, MPI EM) bacteria. The bacteria were selected by Zeocine resistance, analytically DNA digested and sequenced. The bacteria that contained the unrecombined pDONR vector could not produce any colonies, because pDONR vectors carry within theis recombination cassette a gene called ccdB flanked by the recombination sites. This ccdB (control of cell death) gene encodes a protein that is inhibiting the GyrA subunit of DNA gyrase. The ccdB gene was replaced by the respective insert during the recombination. Through the ccdB gene a selective growth of only the correct recombined clones was ensured. Correct recombined contructs carrying the pDONR backbone and the desired insert are desingated 'entry clones' (pENTR). These pENTR serve as a shuttle vector for fast transfer of the insert into any expression vector of interest. The pENTR plamids carry attL sites as result of the recombination between attB and attP sites. In the last recombination, for generating expression constucts (pEXPR), the attL sites recombine with the attR sites of a so-called destination vector (pDEST). This recombination named 'LR reaction' is catalysed by the Gateway LR Clonase II enzyme mix, containing the viral enzymes Integrase and Excisionase, the Integration Host Factor and the reaction buffer. Each LR reaction contained:

pENTR (75-100 ng)	1 µL
pDEST (75-100 ng)	1 µL
LR clonase II	1 µL
H_2O	2 µL
Total reaction volume	5 µL

The reactions were incubated at room temperature for 1 hour or at 4°C overnight. After a sufficient incubation 1 µL of the reaction mix was transformed by electroporation into DH-10b bacteria.

The following sequences were added to the 5'-end of the primers to provide the PCR product with the abbB1 and attB2 recombination sites:

attB1 5'- GGGGACAAGTTTGTACAAAAAAGCAGGCTCC –insert specific sequence – 3'

attB2 5'- GGGGACCACTTTGTACAAGAAAGCTGGGTC –insert specific sequence- 3'

Synthesis of an Oligonucleotide Library by Agilent Technologies

The ordered Oligonucleotide library (Oligo library) has to have a special design. To make it possible to amplify the Oligo library they need to have flanking primer regions on the 5'- and on the 3'-end. In the Oligo inside there is a sequence so called Dec-primer region. To that sequence fits the decoding primer. The name decoding primer because this primer region stands in front of the EXT in 5'-3' direction, and the reverse flanking primer region stands behind the EXT. So the oligo can be decoded with a PCR where the Dec-primer and the reverse flanking primer are used. The *Cis*-element is combined with a TATA-minimal promoter and stands inbetween the 5'-flanking primer region and the Dec-primer region.

The 180mers from the sythesized Oligo library were amplified by Polymerase Chain Reaction (PCR) using the forward and reverse flanking primer. Several established DNA polymerases were tested. To get an acceptable yield and high qualitiy of the DNA we used the HotStartTaqPlus with its proofreading function.

The amplified Oligonucleotides were loaded on a high percentage polyacrylamide gel to get a better size resolution. The corresponding gel band was cut out and the amplfied DNA was extracted. Then the PCR product was used as template DNA for a re-PCR using attB primer for Gateway cloning. These PCR products were also purified over gelelectrophoresis and were subcloned into DONR vectors for cloning into the corresponding EXPR vectors.

2.7 Cell biology Methods

Coating of plastic surfaces with Poly-L-Lysine (PLL)

For a better adherence of low-detaching cells (PC12, 293HEK_FT) culture dishes were pre-treated with 0.02 mg/mL PLL in water. After 1 hour incubation at 37°C the dishes were washed two times with sterile water and dried under the laminar air flow of the tissue culture hood. The coated dishes were stored at 4°C for up to one month.

Passaging of eukaryotic cells

Eukaryotic cell lines (293HEK, 293HEK_FT) were cultivated on 10 cm or 15 cm plastic dishes at 37°C in a humidified incubator containing 5% CO_2. When the cells reached a confluence of 80-90%, cells were subcultured by washing two times with isotone buffer PBS and mild treatment with 1 mL trypsin/EDTA for 1-3 min at 37°C until the cells were detached from the culture dish. With 9 mL fresh growth medium the reaction was stopped and the cells were pelleted by centrifugation for 3 min with 700 rpm. After centrifugation the supernatant was discarded and the cells were resuspended in 10 mL fresh growth medium. For further culturing of the cells 1 mL out of the suspension was plated onto the dishes (coated or uncoated depending on the cell line type).

Thawing of eukaryotic cell lines

Frozen stocks of eukaryotic cell lines were kept in 2 mL cryotubes at -196°C dipped in liquid nitrogen. An aliquot was quickly thawed at 37°C in the waterbath, transferred into a 15 mL falcon tube containing 1 mL pre-warmed fresh growth medium and centrifuged for 3 min at 700 rpm. The supernatant was discarded and the pellet was resuspended in fresh growth medium and plated out onto a 15 cm culture dish.

Generating frozen stocks of eukaryotic cells

To get a good quality of well growing cells the cells were preferentially frozen at a stage of around 60-70% confluence and low number of passages. After a treatment with trypsin the cells were pelleted by centrifugation and resuspended in pre-cooled freezing medium containing a high concentration of serum preferentially Fetal bovine serum (FBS, up to 30%) and DMSO with a final concentration of 5-10% to avoid the formation of water crystals. In each aliquot of frozen cells the density was between 5-8 million cells per 1 mL medium. The tubes were transferred into a freezing box with isopropanol and directly frozen at -80°C. After one day at -80°C the tubes were transferred to -196°C into liquid nitrogen for long term storage.

Transfection of mammalian cell lines

For luciferase assay the cells were typically transfected in 96-well plates. For sequencing experiments and protein analysis by western blotting preferentially cells were transfected in 6-well plates. Depending on the cell line and the well-size, for transfection different cell numbers were plated as indicated below:

Cell line	96-well	6-well	10 cm dish
293HEK	20.000	500.000	3-5 x10^6

WiDr	20.000	500.000	3-5 x10^6
SH-SY5Y	15.000	400.000	
U2OS	15.000	400.000	
MCF-7	15.000	400.000	

For transfection Lipofectamine 2000 Transfection Reagent (Invitrogen, life technologies) was used to introduce plasmid DNA into mammalian cells. Most of the time the cells were transfected in 96-well flat bottom and 6-well cell culture plates (BD Falcon). To achieve optimal results the original protocol of the manufacturer was modified. For transfections mini- as well midi-prep quality DNA was used. Per one well in 96-well plate 10-60 ng of plasmid DNA and 0.2 µL of the Lipofectamine 2000 were diluted in 15 µL Opti-MEM (Gibco, Invitrogen) each. After an incubation of 5 min the two solutions were combined resulting in 30 µL, vortexed for 2 sec and incubated for 20 min at room temperature. After sufficient incubation the growth medium was completely removed from the plate and the cells were covered with 30 µL per well Opti-MEM containing the DNA-Lipofectamine transfection complexes. After an incubation of 2 hours at 37°C, 60 µL of fresh pre-warmed growth medium per well was added to the cells. For transfection in 6-well plates the DNA amount, the Lipofectamine 2000 (LF2000) and Opti-MEM were upscaled. In 6-well plates 300-1000 ng plasmid DNA and 7 µL LF2000 were used in 500 µL Opti-MEM each.

Generation of stable cell lines

To produce stable cell lines expressing the *Cis*-element libraries the plasmids carried a hygromycin B (Hygrogold, Invivogen) resistance gene as selection marker. To have a high complexity 5 million cells on a 10 cm dish were transfected with 10 µg plasmid DNA as described above – for the transfection 40 µL of Lipofectamine 2000 was used. The cells were incubated for 24 hours, then the selection using lethal hygromycin B concentrations for non-transfected was started. Two days after starting the selection the cells were passaged and $1/10^{th}$ was plated on a 15 cm dish for counting the clones and estimating the complexity. After one week selection all cells without hygromycin B resistance were dead. The stable cells were cultivated in large scale for generating of frozen stocks.

Following concentrations of the hygromycin B was used in:

293HEK 75 µg/mL

SH-SY5Y 150 µg/mL

WiDr 400 µg/mL

MCF-7 50µg/mL

Generation of lentiviruses

For some cancer cell lines it is not possible to get foreign plasmid DNA into these cells neither by transfection with Lipofectamine2000 or other phospholipid-particles nor by electroporation of the mammalian cells. For those non-transfectable cell lines we used lentiviral expression systems to infect these cells. To fit in our approach we modified commercial available viral systems as the pLenti6/V5-DEST (Invitrogen, life technologies) and pTripZ (Open biosystems, Thermo Scientific) by classical cloning. In order to get a reproductive-incompetent biological safe virus the viral genes gag, pol and env are not located in one plasmid. We used a two plasmid packaging system with pMD2.G and psPAX2 that are responsible for the correct packaging of the viral RNA. The pLenti6/V5 and pTripZ plasmids carried the viral information of the RNA, in our case the *Cis*-element libraries coupled to EXTs. A virus is only formed in the case that all three plasmids co-transfect into the same cell. For stability reasons of the viral particles we decided to use the pTripZ only. For the production of the viruses we used the 293HEK_FT cell line. The 293HEK_FT cells are optimized cells that stably expresses the SV40 large T-antigen under the control of the human CMV promoter and facilitates optimal production of virus. To generate optimal virus titer the cells had to have a density of 60% on a 15 cm dish. Two hours before transfection the cells got 15 mL of fresh growth medium without any antibiotics (Penicillin, Streptomycin). Then the DNA was mixed in a strict ratio. We used 15 µg of pTripZ with *Cis*-element library (transfer plasmid), 10 µg of psPAX2 and 5µg of pMD2.G diluted into 500 µL of pre-warmed Opti-MEM and mixed by vortexing. To get an impression of the ability to infect cells we added 1 µg of a pTripZ_EYFP construct driven by a CMV promoter. As a transfection reagent we used the polyethylenimine (PEI) that build complexes with the DNA. The ration between DNA and PEI was 1:4 (w/v), so 120 µL PEI was added to the warm DNA Opti-MEM mix and vortexed for 10 sec. The the mix was allowed to incubate for 10 min at room temperature. After incubation the mix was added drop wise to the dish and the cells were incubated overnight at 37°C. The next days (2-3 days) the medium was harvested and stored at 4°C and fresh growth medium without antibiotics was added to the cells. Due to the fact that lentiviruses get secreted out from the cells the harvested medium was purified.

Enrichment of viral particles

To get rid of dead cells and cellular particles the harvested medium was centrifuged in 50 mL falcons for 10 min at 3000 rpm. Then the supernatant was filtered by a 0.45 µm syringe filter (Millex, Millipore). The filtered medium containing viral particles was concentrated using the Ultra-15 Centrifugal Filter Unit (Amicon, Millipore) and centrifugation for 15 min at 3000 rpm. This concentration by centrifugation was repeated until 2 mL of supernatant was left. The viruses were washed two times with 10 mL fresh DMEM and concentrated again. Then the viruses were used directly for infection or stored at -80°C.

Titration of viral particles

Before the viruses were stored 5 µL of virus solution was used to determine the viral titer. To get rid of genomic or plasmid DNA these 5 µL virus solution were treated with DNAse (Promega) at 37°C for 30 min in a 30 µL reaction. To stop the DNAse digest 3 µL of stop solution was added and to open the viral capsids for RNA release the reaction was heated for 10 min at 75°C. Then 4 µL out of this reaction were used for a cDNA synthesis. With a quantitative-PCR reaction the titer of the virus was calculated.

Generating stable cell lines via viral infection

To generate stable cancer cell lines, the cells were infected with a MOI of 10. MOI is the Multiplicity of infection, it is the ratio of infectious virus particles to the number of cells in culture, and for every cell on the dish 10 viruses were added to the medium. The virus libraries do not carry the *Cis*-elements with the EXTs only; they also carry a Puromycin resistance as selection marker. Two days after infection some cells expressed the EYFP construct visible by fluorescent microscopy. After two weeks of selection with Puromycin all cells that were not infected died. The stable cells were cultured and some aliquots were frozen in liquid nitrogen for long term storage.

Luciferase reporter gene assay

For normalization of the measurements of the firefly luciferase (ff-luciferase) activity the readings of a different reporter gene, e.g. renilla luciferase (renilla) were used. Renilla is expressed under a constitutively active promoter. To correct differences in cell numbers, transfection efficiency, RNA and protein expression and general performance of the cells from well to well this type of normalization is used. To get rid of cell performance dependent expression differences we used a combination of three different plasmids to express the renilla luciferase under the control of three different promoters (renilla-Mix). The amount of plasmid DNA was adjusted to balance the expression levels from each of the promoters (SV40:TK:CMV = 10:2:1). The renilla-Mix was co-transfected along with many firefly reporter gene assays. For the measurement of endogenous transcription factor activity by binding on *Cis*-elements and stimulation with drugs only the renilla plasmid with the TK-promoter was used. To monitor the transfection efficiency an equal amount of pEYFPnuc, a CMV-promoter driven expression of the nuclear localized Enhanced Yellow Fluorescent Protein was included (renilla-Mix : pEYFP = 1:1).

Composition of the renilla-Mix:

pRLuc/SV40	100 µg
phRLuc/TK	20 µg
phRLuc/CMV	10 µg
pEYFPnuc	13 µg
10 mM Tris pH 8.5 up to	1.3 mL final conc. 110 ng/µL

Luciferase assays with cell lysis were always performed in 96-weel format. Per well 10-60 ng plasmid DNA was transfected including 10 ng renilla-Mix. To ensure the statistical reliability of the results each assay was performed in 6-12 replicates. This means usually 6-12 wells on the 96-well plate were transfected with the same DNA-Lipofectamine master mix. Depending on the design of the experiment the cells were allowed different time to express the recombinant proteins and reporter genes. After finishing the experiment the growth medium was removed completely and the cells were lysed with 30 µL per well Passive Lysis Buffer (Promega). The plates were incubated for 20 min at room temperature with shaking (200 rpm) and assayed immediately or frozen at -20°C. For the measurement the lysates were transferred into a white plastic microtiter plate to reduce the light signal leakage and cross mixing due to light reflection. The dual luciferase assay measuring the bioluminescence of both firefly and renilla luciferases was done with the help of the Microplate reader Mitras LB940 (Berthold Technologies) and the associated software MicroWin 2000. The Microplate

reader injected to each well 75 µL of the firefly substrate, the reaction was allowed to stabilize for 2 sec and for the next 10 sec the light signals were collected. Then 75 µL of renilla substrate was injected and after 2 sec stabilization time the signals were collected for another 10 sec. Due to a high substrate specificity of the firefly and renilla luciferase during the measurement there is no cross-activation. Moreover the activity of the ff-luciferase is inhibited by the pH conditions of the renilla substrate.

The data were exported from MicroWin2000 and analyzed with the help of Excel (Microsoft). The firefly readings were divided by the corresponding renilla readings producing values in relative luciferase units (RLU). An average over the replicates (6-12 replicates) was taken and the standard deviation was calculated.

Online Luciferase assay measurement

The measurement of luciferase activity can not only be done by a protein assay with cell lysates. We also measured the luciferase activity in living cells. These were stable cell lines containing the pGL4 plasmid with *Cis*-elements coupled to EXTs. The cell lines were split onto 3.5 cm cell culture dishes one day before activity measurement. The luciferase activity was measured for 6 hours with adding 0.1 mM luciferin to normal growth media to get the basal activity. Then the cells were starved for 18 hours in starvation media containing 0.1 % FBS and 0.1 mM luciferin. After starvation the cells were activated with growth media containing 10% FBS and 0.1 mM lucinferin. The activation of the cells was measured for up to 30 hours. For the measurement we used a light-tight incubator coupled to a Hamamatsu photomultiplier tube detector assemblies (Yamazaki et al. 2000, Yoo et al. 2004) to monitor the bioluminescence. With the LumiCycle v 2.0 software from Actimetrics and MS Excel the data were analyzed.

Phenol Chloroform extraction of genomic DNA (gDNA)

An adequate number of stable cells were pelleted by centrifugation. The supernatant was discarded and the cells were lysed by freezing at -20°C for 1 hour, the pellet was thawed and solved in 1 mL DNA extraction buffer. Then 2.5 µL of RNAse A was added and incubated for 3-4 hours at 37°C shaking. After the RNA digest 40 µL of proteinase K was added and the solution was incubated at 55°C overnight. The solution was transferred into a 2 mL Eppendorf reaction tube (Eppi) and the gDNA was extracted by adding 1 mL of Phenol. The sample was mixed by a rotation wheel for 10 min at room temperature. Then the phases were separated by centrifugation for 10 min at 3500 rpm. The upper phase wes transferred into a new 2 mL Eppi and 1 mL Phenol was added. The solution was mixed by rotation wheel for 10 min at room temperature and the phases were seperated by centrifugation for 10 min at 3500 rpm. The upper phase was transferred into a new 2 mL Eppi and 1 mL of a mixture of Phenol/Chloroform/Isoamyalcohol (25/24/1) was added and mixed by rotation for 10 min at room temperature. The phases were separated by centrifugation for 10 min at 3500 rpm again and the upper phase was transferred into a new 2 mL Eppi. Then 1 mL of Chloroform/Isoamylalcohol (24/1) was added and mixed by rotation for 10 min at room temperature. The phases were separated by centrifugation for 10 min at 3500 rpm and the upper phase was transferred into a new 2 mL Eppi. The gDNA was precipitated by adding 1 mL of Isopropanol. The mixture was inverted for 2-3 times and the DNA was pelleted by centrifugation for 30 min at 13000 rpm at 4°C. The pellet was airdried and resuspended in 100 µL of buffer TE.

RNA isolation

To get high quality and yield of RNA we used the RNeasy Mini Kit (Qiagen). Cells stably expressing *Cis*-elements coupled to EXTs were lysed in an appropriate volume of RLT lysis buffer and the lysate was homogenized by pipetting up and down. The lysate was mixed with an equal volume of 70% ethanol vortexed very short and loaded immediately to a RNeasy spin column and cetrifuged for 15 sec at 10.000 rpm. The column was washed with 350 µL buffer RW1 and centrifuged for 15 sec again. To get rid of genomic DNA an on-column DNAse digest (DNAse kit, Qiagen) was performed. Per sample 80 µL of DNAse diluted in buffer RDD was added to the column and incubated for 15 min at room temperature. Then the column was washed with 350 µL buffer RW1 and centrifuged for 15 sec at 10.000 rpm. Then the column was washed two times with buffer RPE dried by centrifugation and the RNA was eluted in 100 µL of RNAse free water.

RNA precipitation

The RNA was precipitated with ammonium acetate salt and ethanol (EtOH). To 100 µL of RNA solution 50 µL of 7.5 M ammonium acetate was added and mixed by vortexing. For a better visualization of the pellet 1 µL of glycogen solution (20 mg/mL, Roche) was added and mixed by vortexing. The RNA was precipitated by adding 450 µL of 100% EtOH. The samples were mixed by vortexing and the RNA was pelleted by centrifugation for 30 min at 13000 rpm. The RNA pellet was washed with 1 mL of 80% EtOH, airdried and resuspended in 5-10 µL RNAse free water.

First Strand cDNA synthesis

First srand cDNA was generated from 0.5-2 µg total RNA using the Superscript III reverse transcriptase (Invitrogen, life technologies). For each sample of an experiment equal amounts of RNA were precipitated and resupended in 4.5 µL RNAse free water, 1 µL of the random nonamer primer (# 4542) was added to the final volume of 5.5 µL. The samples were heated for 2 min at 70°C and placed on ice. Then the other components were added:

5x First Srand buffer	2 µL
0.1 M DTT	1 µL
dNTP mix (10 mM each)	0.5 µL
Superscript III reverse transcriptase (200 U/ µL)	1 µL
Total reaction volume	10 µL

To allow the annealing of random nonamer primers the samples were incubated for 10 min at 25°C. For theRNA dependent DNA synthesis the samples were incubated at 50°C and 55°C each 45min. After the DNA synthesis the samples were heated for 5 min at 85°C for heat inactiovation of the enzyme. 1 µL out of the cDNA synthesis was used for PCR amplification.

Next Generation Sequencing via Ion Torrent PGM

Next Generation Sequencing (NGS) is a general term for new sequencing methods. These methods are faster and cheaper than the classical sequencing by Sanger`s technique. However the main advantage is that sequencing is done in parallel as a high throughput method. Most of the NGS methods are based on a very expensive optic to detect photons. The technique we

used is based on the measurement of protons. Jonathan M. Rothberg (Rothberg et al. 2011) developed this method. In brief, during the elongation of a DNA template a proton is released by the hydrolysis of the dNTP to a dNMP plus the diphosphate. The released proton changes the pH to 0.02 pH units per single base incorporation. This pH shift is measured and converted to a voltage and digitalized by electronics. If the wrong base is in the reaction well no proton is released and the base will be washed out. A signal processing software changes the raw data into sequences in each well. This technique allows the sequencing of the reporter gene libraries TF, Xie and Co with a high complexity.

The EXTs in the gDNA and the synthesized cDNA pools are amplified via PCR using the Dec-primer. A balancer as an internal control and bioinformatic tool is added to the PCR reaction in a dilution of 1:10.000. This balancer mix are unique EXTs with the corresponding primer sites at 3' and 5' end. After this amplification step a so called code-sequence with an adaptor-sequence on its 5' end is added to the amplified EXTs in a second PCR. Each sample gets an unique code to identify the signals out of the sequencing data.

The PCR products of the Code-PCR will be purified. After measurement of the DNA concentration the samples will be diluted. In the next step the samples bind with their adaptor sequence on so called Ion Sphere Paticles (ISPs). These particles will be the template for a next amplification step. After the PCR amplification the Ion Spheres will be purified and enriched. Before every sequencing run the PGM has to be cleaned and initialized. For every run Control Ion Sphere Particles are added as an internal quality control to the enriched ISPs. In the next step the sequencing primers are annealed to the templates. The sequencing chip will be tested on a automatic self-test with the Ion Torrent PGM. After the Chip Check the sequencing polymersae is added to the reaction mix (loaded ISPs, Control ISPs and primers) and incubated for 5 min at RT. Then the sequnecing chip is loaded with the reaction mix by pipetting slowly the complete reaction volume onto the chip without producing air bubbles. Then the chip will be cetrifuged for 30 sec to fill the micro wells with Ion Sphere Particles and the polymerase. In the sequencing run itself the bases are loaded stepwise onto the chip. If the correct base is in the microwell the pH is changed, transformed into voltage and measured. The signals will be transformed into sequences by an automated software from Ion Torrent. The analysis of the raw data will be done with R, a free ware software.

2.8 Biochemical methods

Western Blotting

For the immunological detection of proteins in biological samples we performed western blotting as a classical method (Towbin et al. 1992). In most cases proteins were overexpressed in 293HEK cells by transfection of plasmid DNA carrying the desired protein tagged with a polypeptide for better antibody recognition. Depending on the experiment the cells were lysed between 24 and 48 hours after transfection in different lysis buffers.

Cell lysis

Depending of the experiment and the protein fraction to be analyzed different lysis buffers were used. Lysis buffers differ in their ability to solubilize the proteins, the higher the sodium dodecyl sulfate (SDS) or other ionic detergents the higher the protein yield like RIPA buffer. But these buffers denature proteins. For some applications and antibodies native non-

denatured proteins are needed. There lyse buffers without detergent or with relatively mild non-ionic detergents like Triton X-100 buffer have to be used.

Sodium dodecyl sulfate poly-acrylamide gel electrophoresis (SDS-PAGE)

Depending on the amino acids and their chemical character the charge of proteins vary and they have no linear relationship between weight and charge as DNA molecules have. To make a directed migration of proteins in an electric field possible SDS is used as a detergent and produces a negative charge on the protein surface. The tertiary and secondary structure of the proteins will be destroyed after the heat denaturation step. The polypeptide chains expose the hydrophobic regions and the lipophilic alkyls of the SDS bind via non-covalent bonds at the stoichiometry of 1 SDS per 2 amino acids. An additional treatment with dithiothreitol (DTT) reduces all intra- and intermolecular disulfide bonds.

The protein lysates were mixed with a 4x NuPAGE LDS (lithium dodecyl sulfate) sample loading buffer and DTT with a final concentration of 0.1 M. The samples were heated for 10 min at 70°C. 25-30 µg of protein lysate was loaded into the pockets of the NuPAGE Novex Bis- Tris gels (Invitrogen, life technologies). For the SDS-PAGE we used the XCell SureLock Mini-Cell chamber (Invitrogen) that was cooled by ice and the pre-cooled NuPAGE running buffer. The electrophoresis was performed at 200V for 30 min.

Transfer of proteins to membranes

According to the manufacturer's protocol the transfer of proteins from a SDS-gel to a methanol activated PVDF membrane was performed with the XCell SureLock Western Blot System (Invitrogen) in transfer buffer containing 20% methanol. The transfer was performed at 30V for 2 hours and cooled by ice.

Detection of the proteins

After a successful transfer of the proteins the membrane was incubated in blocking buffer (5% non-fat milk powder in TBS-T for normal antibodies, 5% BSA in TBS-T for α-phospho antibodies) for 1 hour shaking at room temperature. Then the membrane was incubated with the primary antibody in a dilution 1:1000 in blocking buffer for 1-2 hours at room temperature or overnight at 4°C. After the incubation with the primary antibody the membrane was rinsed three times with TBS-T and washed four times in TBS-T shaking for 5 min at room temperature. Then the membrane was incubated with the secondary antibody that is conjugated with horseradish peroxidase (HRP) for 1 hour in a 1:5000 dilution with blocking buffer at room temperature. After a sufficient incubation the membrane was rinsed three times with TBS-T and washed seven times in TBS-T shaking for 7 min at room temperature. This last washing step is the most critical step to get low background signals. For the detection the membrane was incubated for 1 min in enhanced chemiluminescence (ECL) detection solution mix (Perkin-Elmer) by gentle agitation at room temperature. Excess reagent was drained off and the membrane was placed into transparent plastic folders. The visualization of the signals was performed with the ChemoCam Imager (INTAS) a chemiluminescence scanner. The data were analyzed by the software (INTAS).

Co-Immunoprecipitation (Co-IP)

The immunoprecipitation is an often used method to precipitate a peptide-antigen out of a solution via a specific binding of the antigen to an antibody. A Co-IP is used to identify binding partners of that peptid-antigen. When the target protein has a strong binding with its interaction partner, the partner will be pooled out of the solution, too.

In brief, the cell lines were split onto 6-well dishes 18 h before transfection. On the next day the cells were transfected with the needed experiment setup, including the binding and interaction partners and binding controls. 24 hours after transfection the cells were washed with pre-cooled 1x PBS buffer one to two times. All following steps were done on ice to protect the proteins from degradation, especially the phosphorylated ones. Then 600µl of pre-cooled Triton-X cell-lysis-buffer were added to the cells. After the lysis step the samples were sonicated two times for ten seconds to break up the nuclei. To spin down the proteins and get rid of the cell debri a centrifugation step for ten minutes at 13.000 rpm and 4°C followed and the protein concentration was measured. Then the pre-washed and prepared Anti-Flag-M2 beads were added to the samples. The Anti-Flag-M2 beads contain a mouse antibody that recognizes every flag-tagged protein, binds on it and pulls it down. To bind the antibody coupled beads to the flag-tagged protein the samples have to incubate for two hours at 4°C in a rotation wheel. For very strong antibody antigen bonds the incubation is less. Then the antigen-bound beads were purified by several washing and centrifugation steps. The more washing steps you make, the cleaner the western blot is. After the last washing step most of the lysis buffer was removed and the beads were dispensed in 60-70 µl of loading buffer containing DTT. Then the samples were heated for 10 minutes at 70°C, as protein denaturation step. The samples were load on SDS-gel depending on the protein amount and western blotting for protein identification followed.

3. Results

3.1 Construct design of the TF, Xie and Co reporter gene libraries

The *Cis*-element coupled EXT oligonucleotides were designed for the measurement of transcription factor activity. The aim of EXTassays was to generate a valid high-throughput method for the readout of TF activity and signalling pathways. The EXTs are coupled to regulatory elements. All libraries were built according to the same scheme (figure 3.1); an EXT is 5' and 3' flanked of a library specific sequence for amplification and cloning. On the 5' end the 36 bases long *Cis*-element with a 41 nucleotide stretch of a minimal promoter (TATA) as transcription regulating region is added. On the 5' end of the *Cis*-element an 18 base pair long library specific primer is added for amplification and cloning. Each *Cis*-element is coupled to a unique EXT containing three technical replicates.

Three different libraries were generated, each library contained specific features for solving different issues. The first library, the TF library is based on the TRANSFAC database. There all data on eukaryotic transcription factors with the experimentaly-proven *Cis*-elements are provided. The TF library contains all experimental validated human and mouse *Cis*-elements out of the TRANSFAC database of the release 12.1 (2008.1) (Wingender et al.1997). The TF library contains in total 3987 constructs.

The second library is based on the comparative analysis of the genomes of humans, rats, mice and dogs (Xie et al. 2005). The Xie library consists phylogenetic conserved regulatory elements out of promoter regions and 3' untranslated regions (UTR) from rat, mouse and human that may work as transcription factor binding sites (Xie et al. 2005). The Xie library contains 2223 constructs in total.

The third library, the Co library is based on the COMPEL database. In this database combinatorial motifs with specific transcription factor-DNA binding / interaction are enclosed. The Co library contains human and mouse composite regulatory elements from COMPEL database of the release 12.1 (2008.1) (Heinemeyer et al. 1998). The Co library contains in total 948 different constructs.

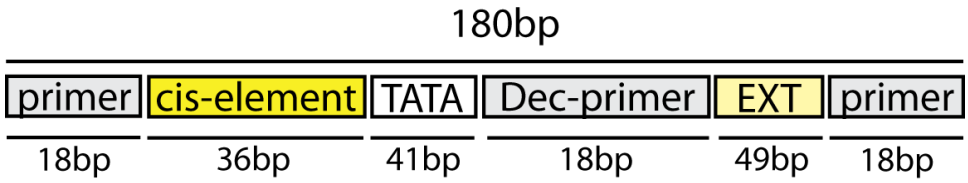


Figure 3.1: construct design of the oligonucleotides
The *Cis*-element is coupled 3' to a minimal promoter region (TATA-box) and 5' and 3' flanked with primers. On the 3' end the EXT and the amplification primer is added.

3.1.1 Covering the complexity of the libraries

To reduce the risk of any loss of complexity during the cloning of the libraries every cloning and subcloning step was upscaled to reach a 10 fold complexity at least.

This means for the TF library more than 40.000 bacterial clones in every subcloning step. For the Xie library more than 22.000 and for the Co library more than 10.000 bacterial clones.

3.2 DNA Quality control

Due to the fact that the synthesis of the three oligonucleotide libraries was done with a novel technique - the on-chip synthesis with agilent technologies – random samples for quality control had to be generated. After the amplification via PCR using the HotStartTaqPlus Polymerase the oligonucleotides were subcloned into pDONR-Zeo to generate the ENTR-libraries. These ENTR-libraries were subcloned via Gateway-cloning into the pGL4 destination vector. Out of these EXPR libraries bacterial single clones were picked out as random samples. The single clone plasmid DNA was extracted and sequenced. In total 122 single clones were picked and sequenced.

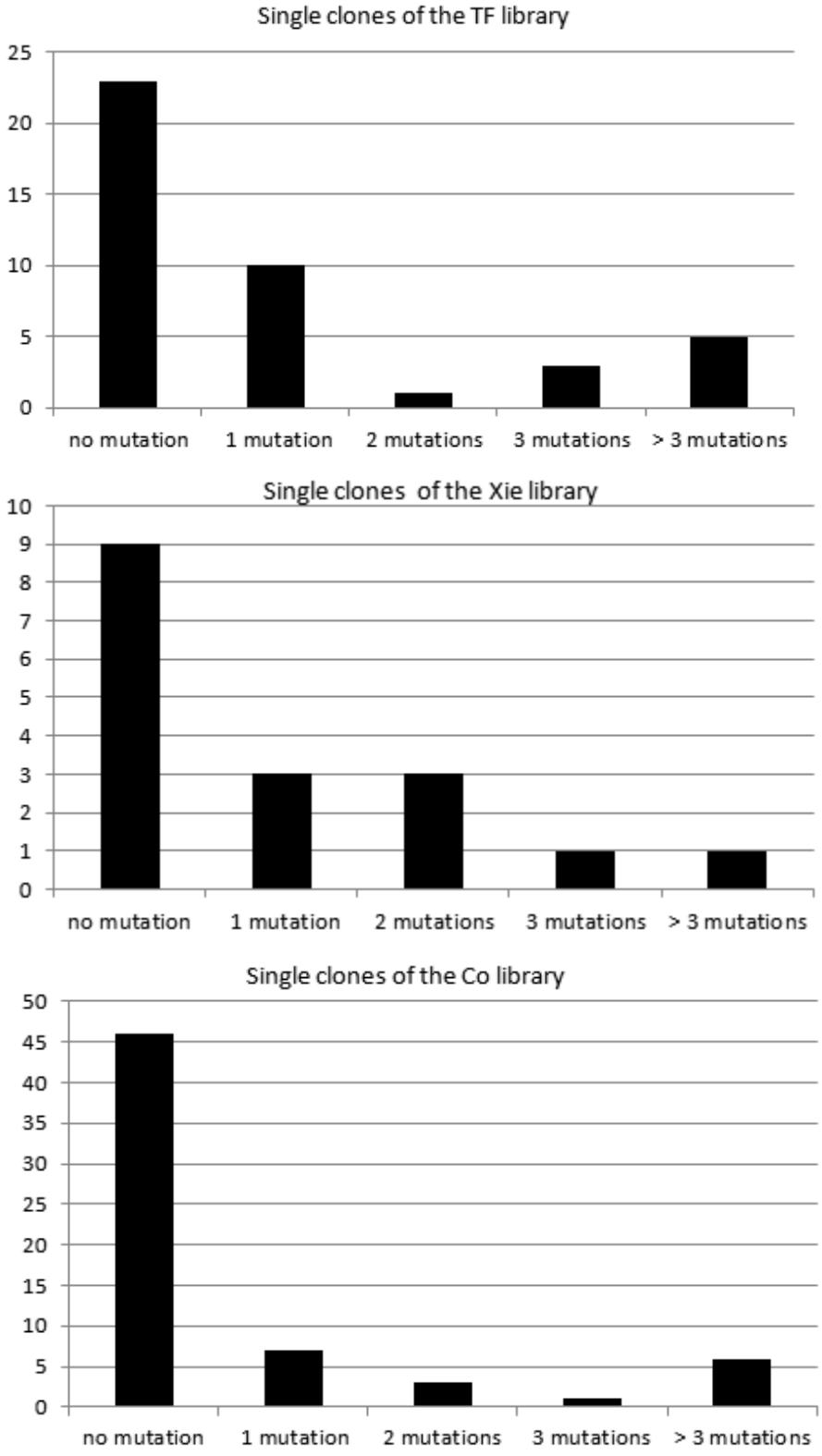


Figure 3.2: Sequencing results of the random picked samples
In the TF library 55% of the picked random samples had a 100% sequence match, the samples of the Xie library had to 52% a 100% sequence match. And 73% of the Co samples had a 100% sequence match. Out of all samples 80% should be functional.

Out of 42 picked single clones of the TF library 55% (23 clones) have a 100% sequence match. 23 % (10 clones) show one mismatch mutation and 11 % (5 clones) have more than 3

mutations within the oligo structure (figure 3.2). From the Xie library 17 single clones were picked and sequenced. 52 % (9 clones) were free of mutations, 17 % (3 clones) had one mutation and 5 % (1 clone) showed more than three mutations (figure 3.2). Within the Co library 73 % (46 clones) out of 63 picked single clones showed a 100 % sequence match. 11 % (7 clones) had one (a single) point-mutation and 9 % (6 clones) of the Co library showed more than 3 mutations (figure 3.2).

Matching the sequences does not say anything about the functionality. Functionality depends in which part of the oligonucleotide the mutations exist. To know more about the functionality we performed a deep sequencing of the plasmid DNA of all three libraries within the pGL4 expression vector.

In the analysis of the deep sequencing 10 % of all constructs were coupled to an EXT that belong to another library. These could be filtered out within following experiments via bioinformatic computing. 78 % of the EXTs of the TF library could be detected within the corresponding library. In the Co library 77 % of all EXTs that belong to the Co library could be detected. From the Xie library only 53 % could be detected via deep sequencing (figure 3.3), the other EXTs of the Xie library were not functional.

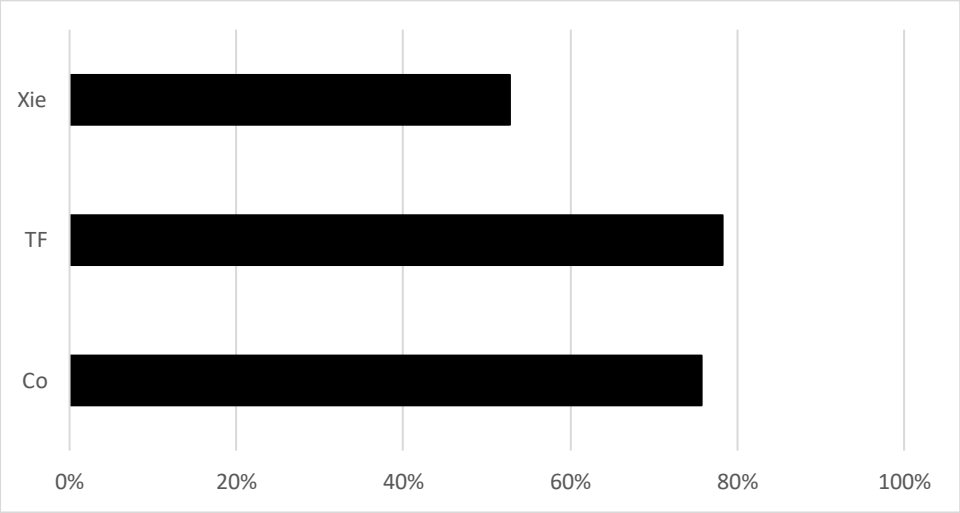


Figure 3.3: Results of the library wide deep sequencing
Only 53% of the EXTs of the Xie library could be mapped within the Xie library. Nearly 80% of the EXTs of the TF and Co library could be mapped in the deep sequencing.

3.3 Proof-of-principle experiments

After testing the functionality of the EXTs via deep sequencing, the libraries had to be tested in *Cis*-element functionality. Selected single clones out of the libraries in the pGL4 (promega) expression vector were used for this approach.

In contrast to the pGL3 the pGL4 expression vector is characterised with a codon optimization especially for usage in mammal cell lines, the firefly luciferase2 (luc2) an optimized luciferase with a reduced background activity, removal of cryptic transcription factor binding sites and a hygromycine resistance(HygR) in addition, too. In addition a rabbit β-Globin intron was cloned in 5' position to the luc2 with classical cloning method (figure 3.4).

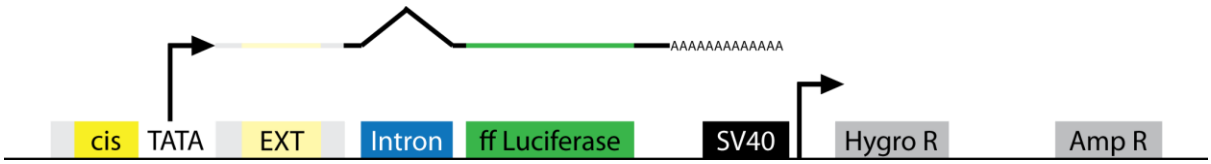


Figure 3.4: scheme of the pGL4-Intron plasmid
The oligonucleotide is 5' of an intron of the rabbit ß-globin gene and the firefly luciferase 2. The transcript of the plasmid is the EXT with its flanking primers 5' and 3', the intron that will be spliced out and the firefly luciferase 2.

By using the selected single clones classical luciferase assays were performed. In these luciferase assays a transient co-transfection of the *Cis*-element and the corresponding transcription factor was done in 293 HEK wt cells. The used *Cis*-elements were the canonical E-Box, the cAMP-responding element (cre), the nuclear factor kappa B binding site (NFκB) and the hypoxia induced factor (HIF) binding site.

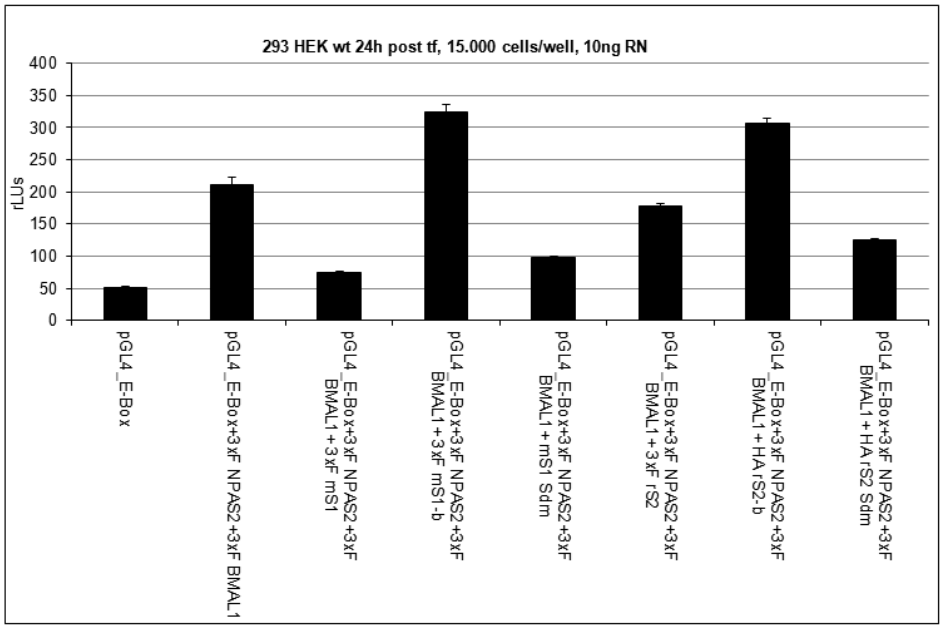


Figure 3.5: single construct validation:
With the activation of the transcription the firefly luciferase values and the relative luciferase units (rLUs) increase. Sharp 1 and 2 (S1 and S2) work as repressors of the transcriptional machinery; the rLUs decrease compared to the sample with NPAS2 and BMAL1.

The canonical enhancer box (E-Box) has the palindromic sequence CACGTG and represents a binding motif for transcription factors to initiate transcription. E-boxes represent *Cis*-elements for neuronal PAS domain protein 2 (NPAS2) and the brain and muscle Arnt-like protein 1 (BMAL1). Both transcription factors belong to the basic Helix-Loop-Helix –PAS-containing family. NPAS2 and BMAL1 are able to form a functional heterodimer that is able to bind to the E-Box, leading to transcriptional initiation. E-Boxes with NPAS2/BMAL1 or CLOCK/BMAL1 binding are the main components of the molecular clock (Ko and Takahashi 2006). The two basic Helix-Loop-Helix (bHLH) transcription factors SHARP1 (DEC2, S1) and SHARP2 (DEC1, S2) work as negative regulators of the molecular clock.

In all classical luciferase assays 10 ng of renilla-Mix plasmid DNA were co-transfected for data normalisation reasons. In the cell culture experiment NPAS2 and BMAL1 bind as transcription factors on the canonical E-Box and lead to transcription of the luc2 gene and

further expression of the luc2 protein. The assay show a significant increase of the relative luciferase units (rLUs) compared to the baseline (figure 3.5). S1 and S2 work as repressors; the transcription is downregulated. The assay shows a significant reduction of the rLUs in the S1 and S2 co-transfection samples. S1 works as a more efficient repressor compared to S2. In the S1 and S2 bHLH transcription factors the basic domain interacts with the DNA. In this assay functional mutations of S1 and S2 were also tested (figure 3.5).

Loss-of-function mutations of these two repressors have less influence in the transcriptional regulation like the –b variant where the basic domain is missing and the protein is not able to bind on the DNA or the site directed mutagenesis (Sdm) where a point mutation leads to a weaker protein DNA binding (Rossner et al. 2008).

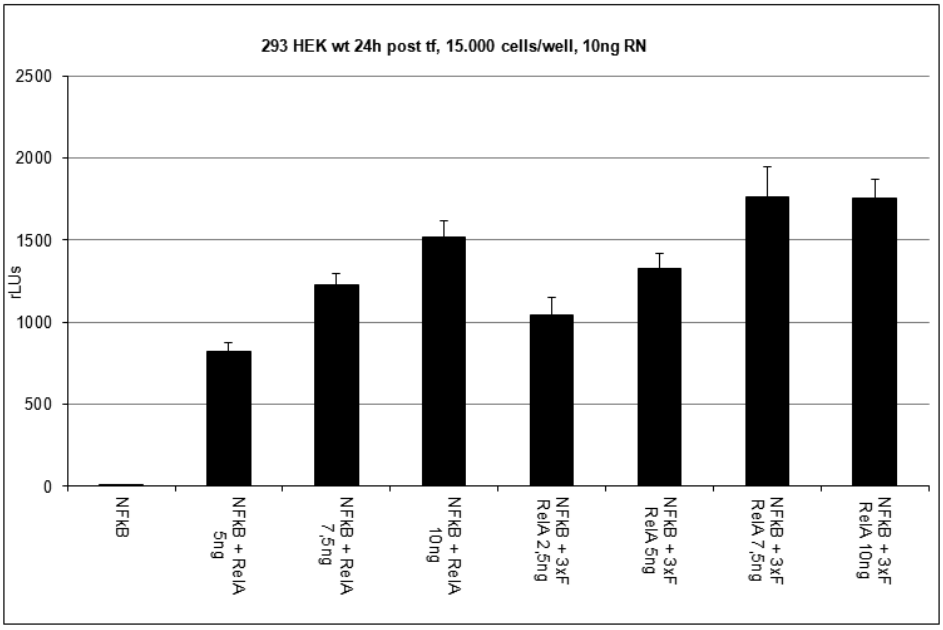


Figure 3.6: single clone validation of the NFκB *Cis*-element:
Rel-a the binding partner of the NFκB complex binds on the *Cis*-element and activates the transcription. The luciferase activity increases depending on the co-transfected Rel-a amount.

The next performed luciferase assay was focused on the binding site of the NFκB complex. The nuclear factor kappa-light-chain-enhancer of activated B cells (NFκB) is a protein complex acting as a transcription factor. NFκB plays a pivotal role in cell differentiation, proliferation, inflammation and cell survival (Oeckinghaus and Ghosh 2009). The NFκB protein complex is expressed in almost every cell type. The NFκB / Rel transcription factor family is characterised with the Rel homology domain (RHD). In mammals, also in humans five different Rel proteins are existing containing RELA (alias p65), RELB, c-REL, NF-κB1 and NF-κB2. The acting transcription factor NFκB is a heterodimer of NF-κB1 and RELA. RELA and NF-κB1 itself are able to form homodimers and bind to the DNA. Whereas the thermodynamically preferred form is the heterodimer (Chytil and Verdine 1996).

In this assay two different plasmid backbones containing murine *Rel-A* were co-transfected. Rel-A is forming a homodimer or heterodimer with endogenous NF-κB1. These transcription factor complexes bind directly on our *Cis*-element called NFκB (Brivanlou and Darnell 2002) and initiate the transcription of the *luc2*. In the assay the rLUs increase in proportion to the co-transfected *Rel-A* containing plasmid.

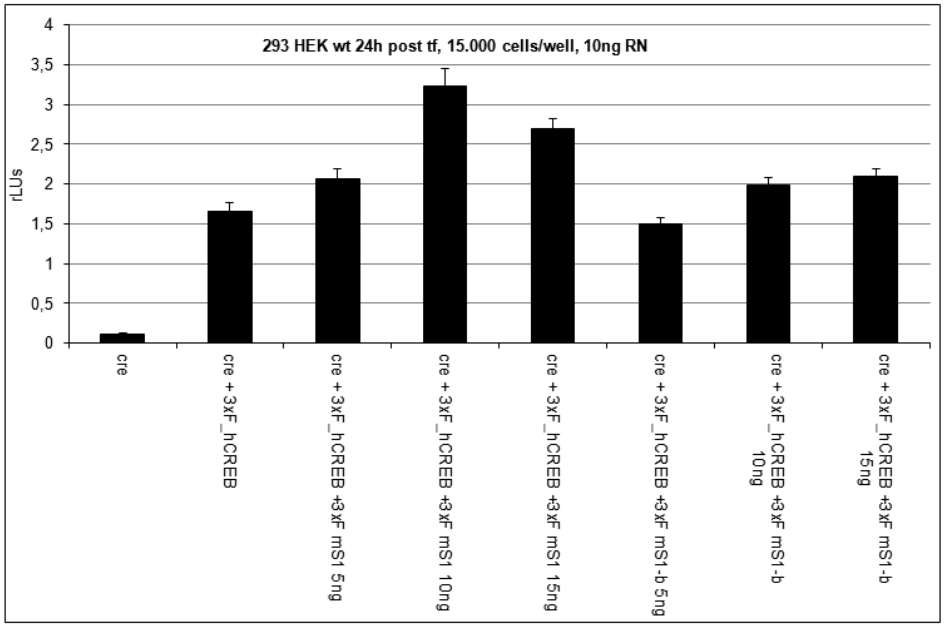


Figure 3.7: single clone validation CRE *Cis*-element: The activated co-transfected CREB forms a dimer and binds on the CRE *Cis*-element; the firefly luciferase activity is upregulated.

The cAMP response element binding protein (CREB) is a transcription factor that belongs to the CREB / ATF transcription factor family. The characteristic domain of these transcription factors is the homological basic leucine zipper (bZIP) domain (Lee and Masson 1993). CREB is activated by phosphorylation of various proteins that act as kinases as PKA (protein kinase A), MAPK, Ca^{2+} - calmodulin-dependent protein kinase IV (CaMKIV) (Lee and Masson 1993). Activated CREB is able to form homo- or heterodimer and bind to the DNA on its specific *Cis*-element. CRE, the cAMP response element with its consensus core sequence TGACGTCA is found in many enhancer or promoter regions. In the assay the pGL4 plasmid containing the CRE *Cis*-element was transfected into 293HEKwt cells (baseline) (figure 3.7). In the activation assay 10 ng of *CREB* containing plasmid was co-transfected. The phosphorylated CREB is able to bind on the CRE Cis-element and transcription of *luc2* is enabled. In the co-transfection assay the rLUs are increased compared to the baseline (figure 3.7).

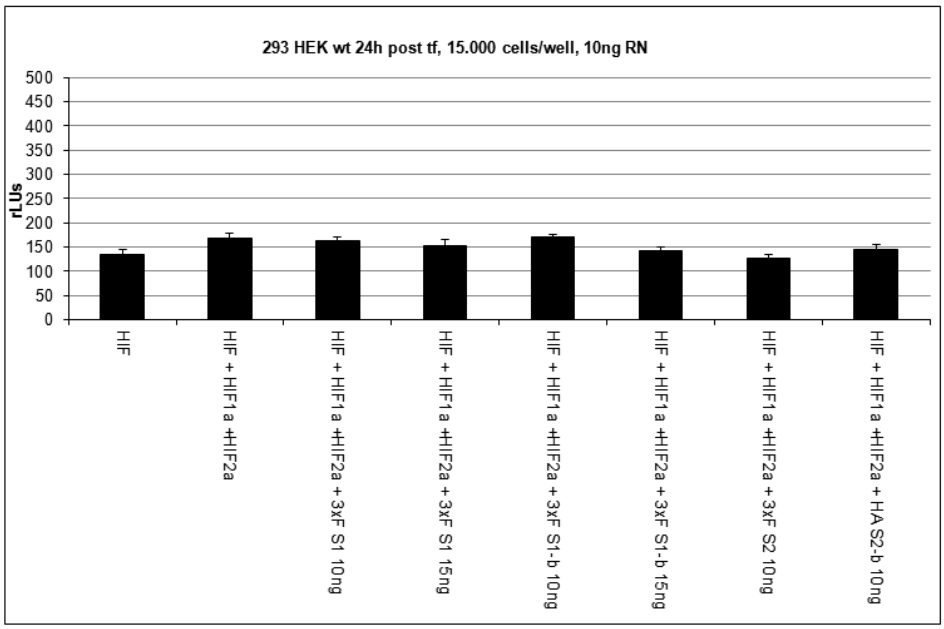


Figure 3.8: single clone validation HIF *Cis*-element: The co-transfected HIF1a and HIF2a are not able to form a functional heterodimer; binging on the HIF *Cis*-element does not occur; the firefly luciferase activity is not upregulated.

The hypoxia inducible factor 1 (HIF1) is a transcription factor that belongs to the bHLH transcription factor family (Semenza et al. 1997). It acts as a transcription factor of genes including erypoetin, vascular endothelial growth factor (VEGF) and lactacte dehydrogenase A (LDHA) (Semenza et a. 1997). The acting transcription factor HIF 1 is composed of two different proteins, HIF1α and aryl hydrocarbon receptor nuclear translocator (ARNT), also known as HIF1β. During hypoxia in cells HIF1α and ARNT form a heterodimer, HIF1. HIF1 is able to bind with the basic domain on its *Cis*-element HIF with its core sequence RCGTG (Semenza et al. 1997). In this assay the specificity of the HIF *Cis*-element was tested. The cell culture experiments were performed under hypoxic conditions with only 5 % oxygen during the cell culturing. The transfected pGL4 plasmid containing the HIF *Cis*-element as key motif. During hypoxia in the cell endogenous activated HIF1α and ARNT form the working heterodimer HIF1 and bind on the HIF *Cis*-element. The control shows during hypoxia a relatively high baseline (approx.. 150 rLUs). The co-transfected HIF1α and HIF2α are not able to form an acting heterodimer and to initiate the transcription of *luc2*, the rLUs are not elevated compared to the baseline (figure 3.8).

All single clone validation assays were done with three biological replicates at least.

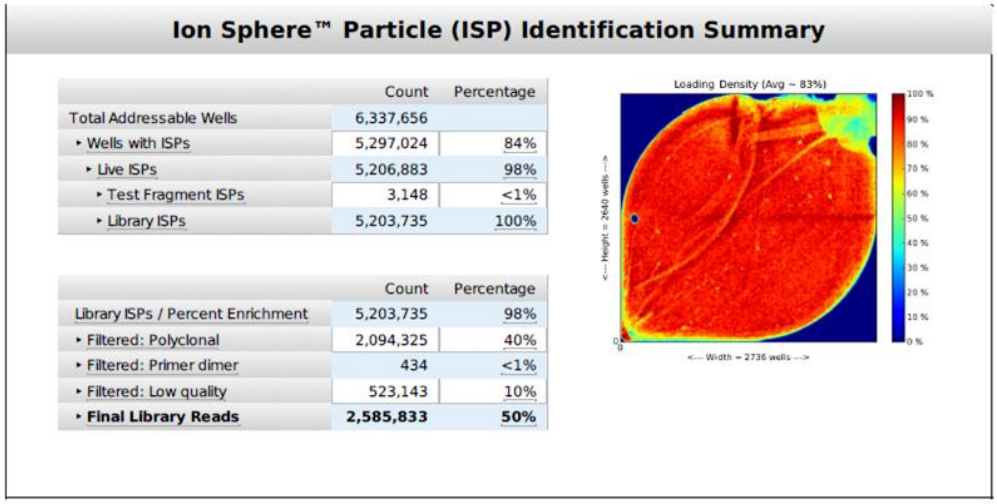
3.4 Transient cell culture experiments

After the proof-of-principle experiments the first transient *Cis*-element library assays were performed. In a series of experiments proliferation assays were performed but within the measurement with the NimbleGen hybridization method technical problems occurred and the measurement failed (figure 3.9).

Figure 3.9: Scan of a NimbleGen chip after hybridization
Due to technical problems the hybridization with the NimbleGen chips didn't work, here a chip after the hybridization. The visible parts on the chip are the technical direction controls and dust.

In the next step the validation of the proof-of-principle experiments with the EXT reads via Next Generation Sequencing was performed. In a co-transfection experiment as shown with the single clones the whole libraries were transiently transfected, together with the plasmid carrying the *Rel-A* gene in PC12 tetOFF rat cell line. After an incubation of 24 hours the cells were lysed and the RNA and DNA were extracted. In a c-DNA synthesis step the RNA samples were reverse transcribed and both the RNA and DNA samples were amplified and re-amplified with the corresponding sequencing primers in PCR steps. These PCR products were purified and sequenced via Next Generation Sequencing.

More than 2.5 million reads were detected (figure 3.10), so technically the experiment worked. In the analysis a high variability of the EXT signals out of the DNA and RNA samples especially the DNA samples was observed (figure 3.11). It was impossible to make precise clusters of the activity of transcription factors within cells with these imprecise data.



Ion Sphere™ Particle (ISP) Identification Summary

	Count	Percentage
Total Addressable Wells	6,337,656	
▸ Wells with ISPs	5,297,024	84%
▸ Live ISPs	5,206,883	98%
▸ Test Fragment ISPs	3,148	<1%
▸ Library ISPs	5,203,735	100%

	Count	Percentage
Library ISPs / Percent Enrichment	5,203,735	98%
▸ Filtered: Polyclonal	2,094,325	40%
▸ Filtered: Primer dimer	434	<1%
▸ Filtered: Low quality	523,143	10%
▸ **Final Library Reads**	**2,585,833**	**50%**

Figure 3.10: sequencing quality report:
With a loading density of 100 % the sequencing run worked technically fine. Overall more than 2.5 M reads could be detected and analyzed.

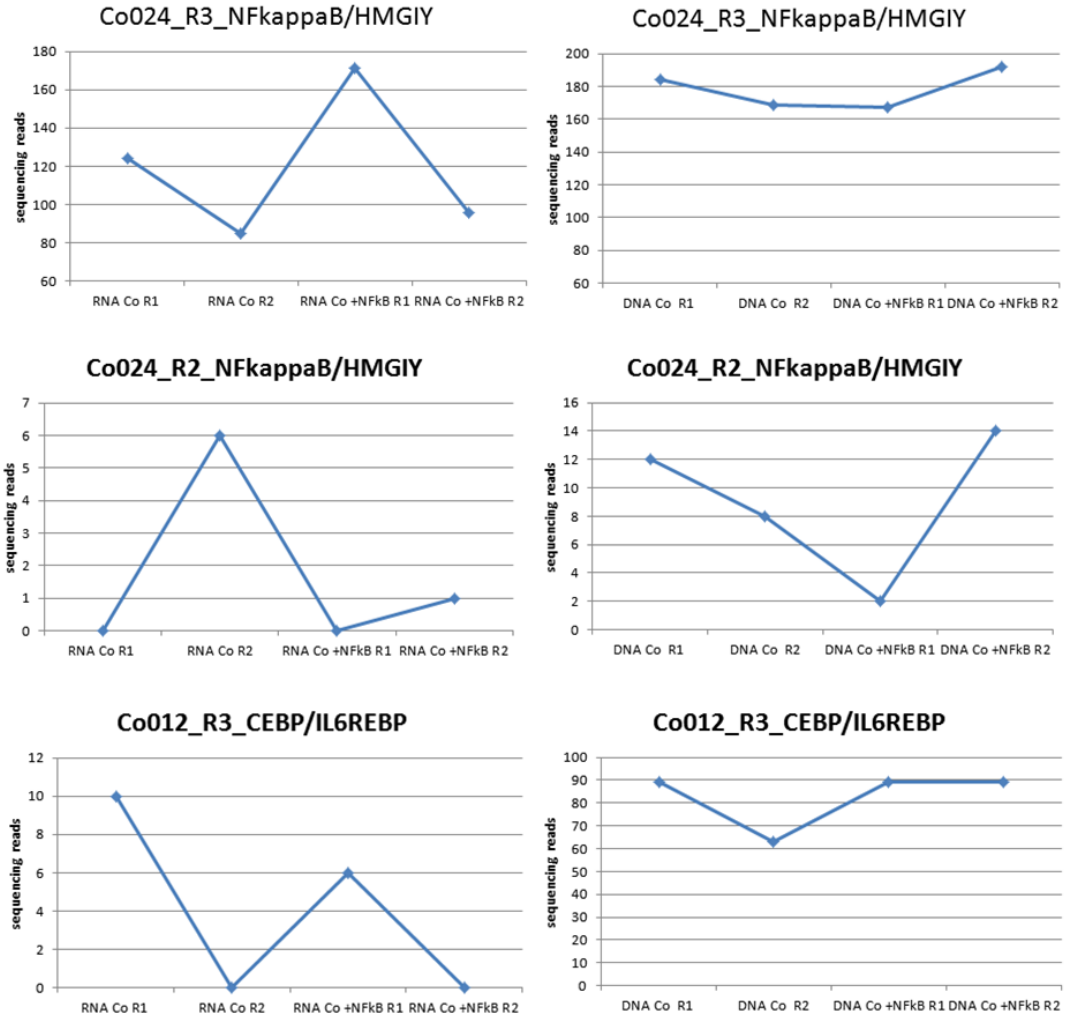


Figure 3.11: sequencing reads of the transient Rel-a Co-transfection experiment:
Examples of DNA and RNA reads, with this variety within the data especially within the DNA samples, it is not possible to make a correct statement about the activity of the transcription factors in the living cells.

3.5 Stable cell culture experiments

In the previous experiments was shown, that these complex *Cis*-element libraries are inappropriate for transient in vitro experiments and signal read outs on RNA level. With the hygromycin resistance gene, the optimized pGL4 plasmid containing the *Cis*-element libraries was used to create stable cell lines. The generation of stable cell lines was performed as described in chapter material and methods 2.7. For all cell culture experiments stable human cancer cell lines with a low passage number were used. Within these cells the pGL4 plasmid was transfected and the cells were selected with an appropriate concentration of hygromycin B tested on previous killing curves (material and methods 2.7).

For the cell line WiDr, a colorectal cancer cell line coming from the colorectal cancer cell line HT-29 6000 stable single clones were created. For the cell line 293 HEK wt 5000 stable single clones were created. SH-SY5Y a human neuroblastoma cell line were used for the generation of 1000 stable single clones and the human cell line of an adenocarcinoma of the breast MCF-7 was also used for 1000 stable single clones.

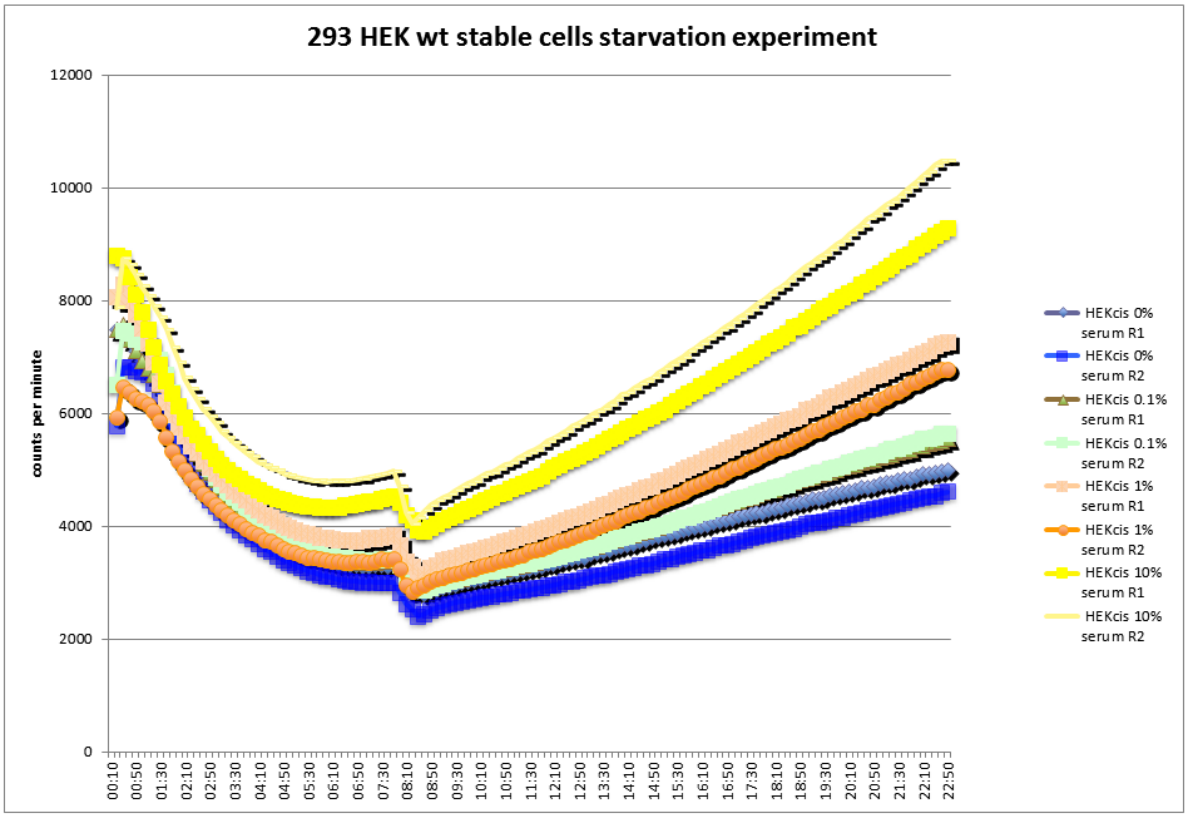


Figure 3.12: online luciferase measurement during starvation:
The figure shows the luciferase values during 24 h of starvation with duplicates. The replicates within the subgroups show a high similarity and a clear trend.

The first planned experiment with the conventional stable cell lines was a proliferation assay with previous serum starvation. Therefore different serum starvation conditions were tested by an online luciferase The best starving conditions for a planned proliferation assay were tested by an online luciferase measurement for 24 hours. In this test the starvation with no fetal bovine serum showed the lowest luciferase activity (figure 3.12) but these cells showed also the highest apoptosis rate (rounded cells) (figure 3.13. The cells with starvation condition with 0,1% fetal bovine serum showed much more vital cells with low luciferase expression (figure 3.13). Therefore for all following proliferation assays 0.1% FBS as

starving condition was used. This condition showed the best performance of the cells as well in the luciferase measurement as histologically compared to the 10% FBS control.

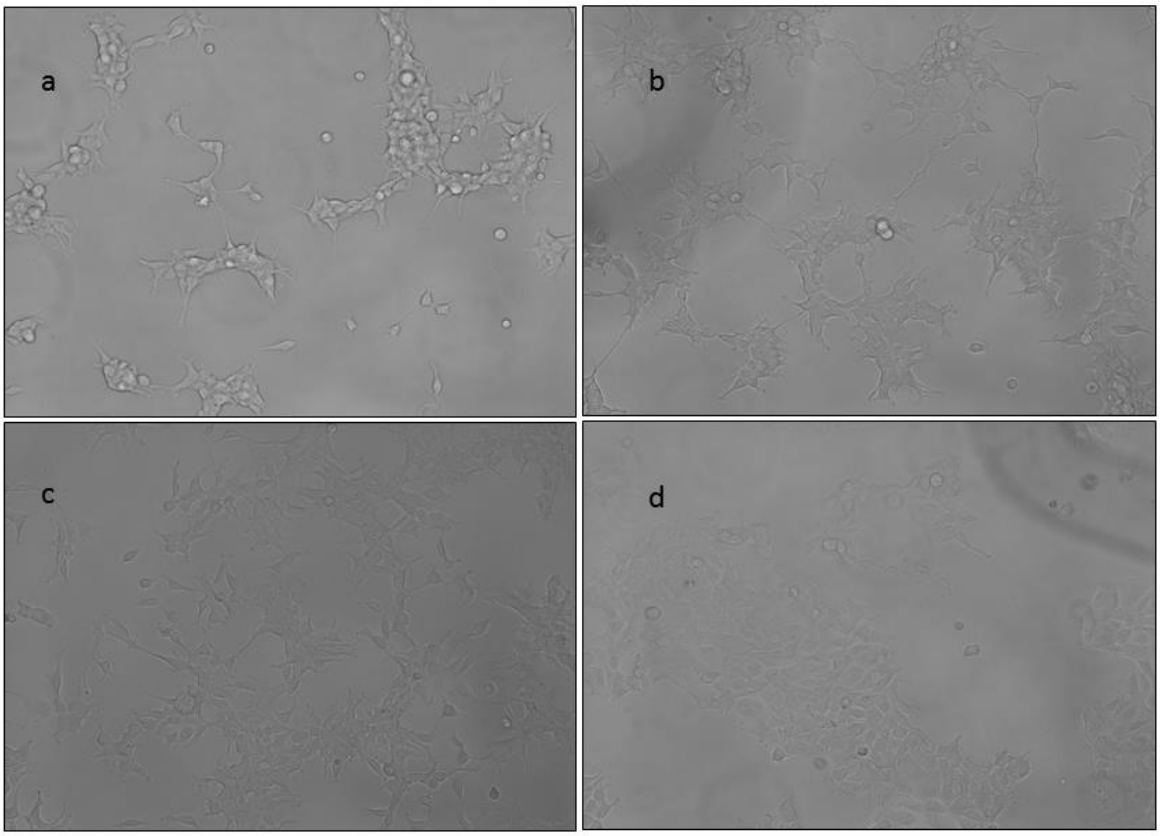


Figure 3.13: 293 HEK wt stable cells after 24 h of starvation and online luciferase measurement:
293 HEK cells after 24h starvation with 0% FBS(a), 0.1% FBS(b), 1% FBS(c) and 10% FBS as control (d). Some cells start apoptosis in the samples with no or low concentrations of FBS (cells in round figures), however the cells with 0.1% FBS look much more alive than the cells without FBS. In between the samples with 1% and 10% FBS (control) no visible effect was demarcated.

The first starvation and proliferation assays were performed in the stable cell line SH-SY5Y. This assay was performed with three biological replicates for each condition. The samples were starved in culture media containing 0.1% FBS for at least 18 hours. The control samples were starved for further 48 hours while the proliferation samples were stimulated with 10% FBS and PMA (10 ng/l). PMA acts as a direct activator of the protein kinase C (Rydholm et al. 1995)

Within the starvation and activation phase the luciferase activity was determined via the online luciferase reader Lumicycle. The stimulation group showed elevated luciferase activity compared to the starvation group (figure 3.14).

The samples were lysed into RLT buffer and the DNA and RNA were extracted. The RNA samples were reverse transcribed in a cDNA synthesis step; all samples were amplified via PCR and sequenced in an Ion PGM (life technologies) (material and methods 2.6). The sequencing results were normed to total reads and 155 different EXTs of the TF library corresponding to their *Cis*-element could be identified. Within these *Cis*-elements e.g. AP1, SREBP, E-Box, GATA and IK could be found (table 3.1).

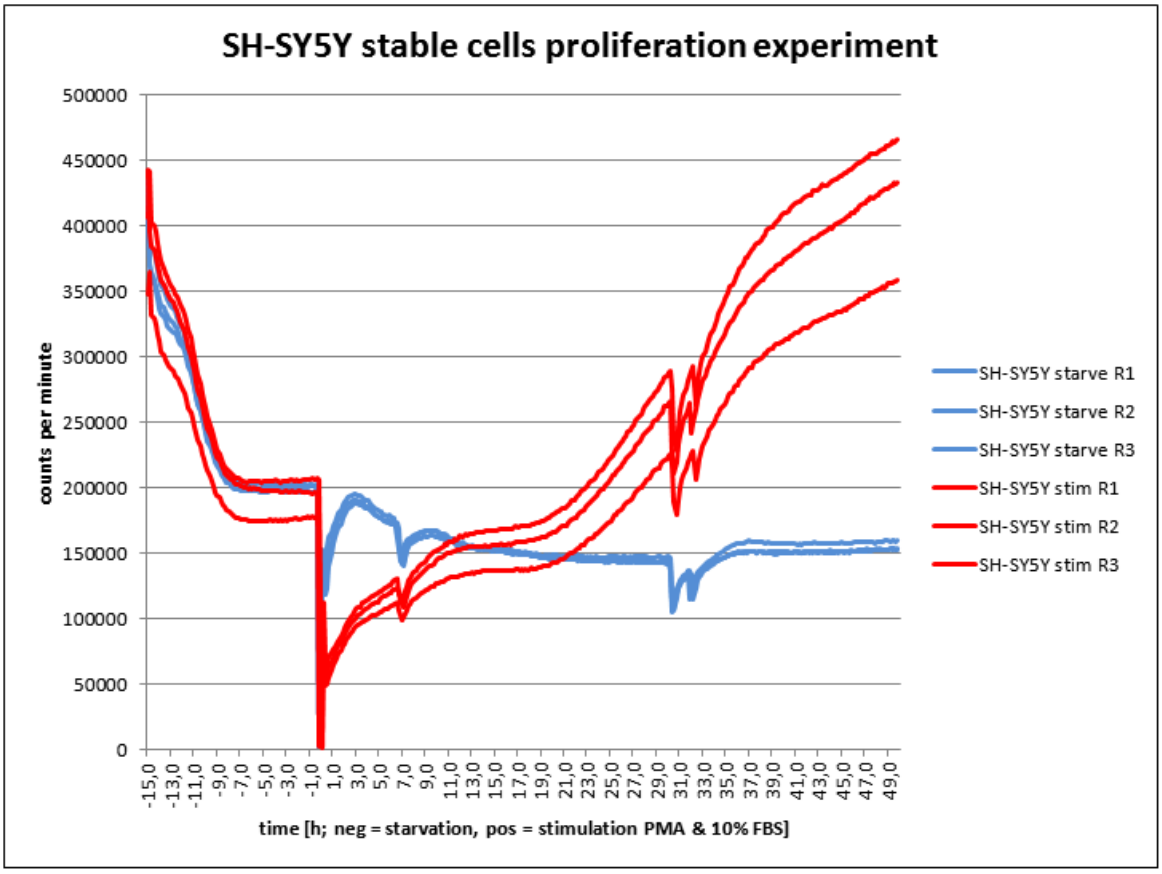


Figure 3.14: online luciferase measurement:
The stable SH-SY5Y cells were starved with media containing 0.1% FBS and then stimulated with 10% FBS and PMA. The stimulation group showed higher luciferase activity during the stimulation phase.

In this experiment in the SH-SY5Y cells upregulated EXT expression during the FBS stimulation was shown of the *Cis*-elements IK, SREBP, CRX, AP1 (table 3.1). IK_Q5 showed a 300 times upregulation, AP1_Q2_01 and NKX25_Q5 showed doubled activation as well as GATA_Q6 and CRX_Q4. A downregulation of the transcription factor activity during the stimulation phase was seen by MTF1, CEBP, NFAT and MEF2 e.g. (figure 3.15).

Cis-element	SHSY5Y starv. R1	SHSY5Y starv. R2	SHSY5Y starv. R3	SHSY5Y PMA R1	SHSY5Y PMA R2	SHSY5Y PMA R3	ratio mean stim/starv
IK_Q5	0	0	0	426	287	202	305.0
SREBP_Q6	0	0	0	17	53	22	30.7
CRX_Q4	35	0	0	209	44	34	8.2
AP1_Q2_01	0	20	30	53	42	107	4.0
TFE_Q6	28	0	44	63	78	95	3.3
NKX25_Q5	404	164	300	626	720	1005	2.7
AP1_Q2_01	276	257	438	863	856	891	2.7
CREB_Q3	151	156	57	205	469	211	2.4
CRX_Q4	64	90	172	244	179	275	2.1
CRX_Q4	685	771	1428	1961	1606	2579	2.1
GATA_Q6	612	523	293	1008	1135	866	2.1

Table 3.1: abstract of the sequencing results of the stable SH-SY5Y proliferation experiment

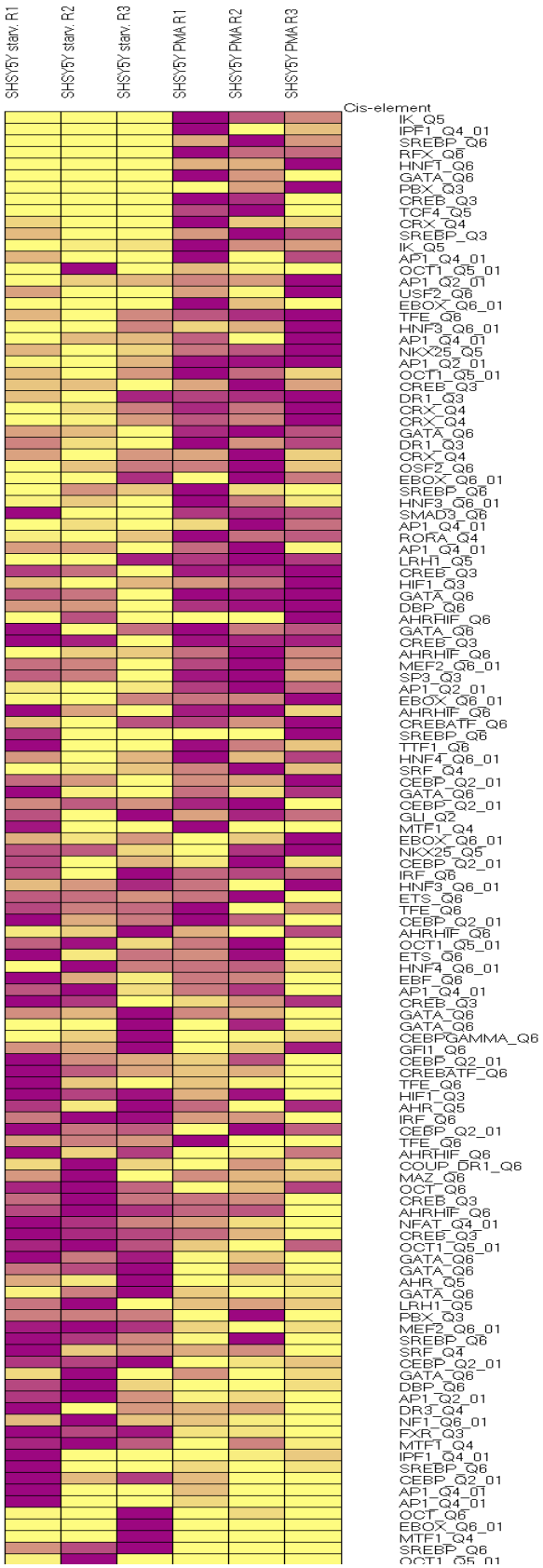


Figure 3.15: heatmap with the EXT reads of the proliferation experiment with stable SH-SY5Y cells

The next performed proliferation assay was done with the conventional stable cell line 293 HEK. There the same conditions as described in the SH-SY5Y cells were used with four biological replicates for each condition. The cells were starved for 18 hours and stimulated with 10% FBS and PMA for 26 hours, lysed, processed and the following samples sequenced with the Ion torrent PGM.

As showed as in the previous experiment the luciferase reads of the stimulated samples were significantly elevated compared to the starvation samples in luciferase measurement (figure 3.16).

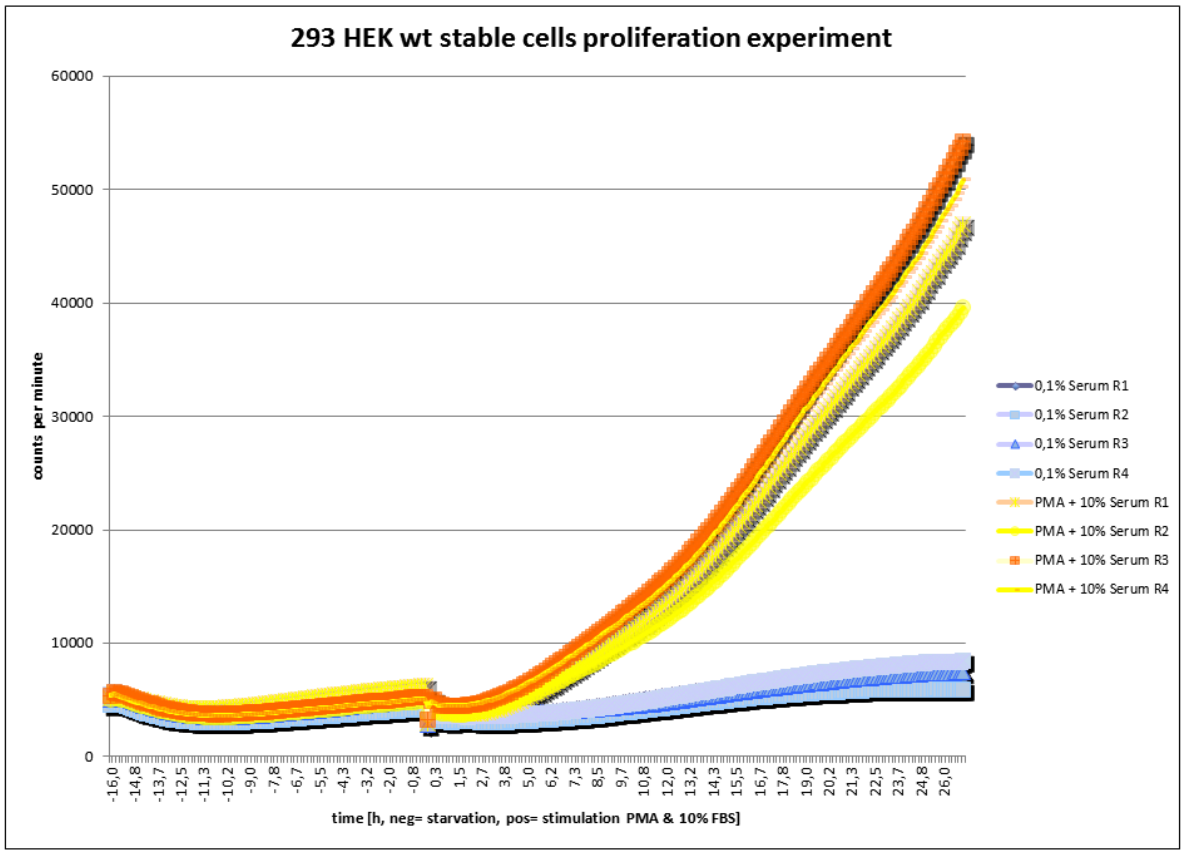


Figure 3.16: online luciferase measurement:
The stable 293 HEK wt cells were starved with media containing 0.1% FBS and then stimulated with 10% FBS and PMA. The stimulation group showed higher luciferase activity during the stimulation phase.

After sequencing of the samples via Next Generation Sequencing (NGS) 266 different EXTs of the TF library could be identified and were normalized to total reads. Here several AP1 *Cis*-elements showed upregulated EXT expression (table 3.2). Within the first 10 *Cis*-elements AP1 binding sites appeared five times and were upregulated up to 200 times (table 3.2 and figure 3.17).

Cis-element	HEK starv R1	HEK starv R2	HEK starv R3	HEK starv R4	HEK PMA R1	HEK PMA R2	HEK PMA R3	HEK PMA R4	Ratio mean stim/starv
AP1_Q4_01	1	2	1	1	185	236	175	598	238.8
ETS_Q6	0	1	0	0	3	8	4	11	26.0
MAF_Q6_01	0	0	23	0	99	37	386	0	22.7
PPAR_DR1_Q2	53	24	67	44	769	988	732	1160	19.4

AP1_Q4_01	0	6	0	0	5	36	44	31	19.3
TCF4_Q5	471	339	351	277	5330	6598	5513	6317	16.5
AP1_Q2_01	69	45	21	44	622	774	612	868	16.1
AP1_Q2_01	72	102	64	107	1052	1474	1139	1446	14.8
AP1_Q4_01	6334	6017	4606	5265	79437	89262	66437	82067	14.3

Table 3.2: part I of the sequencing results of the TF library RNA stable 293 HEK wt proliferation experiment

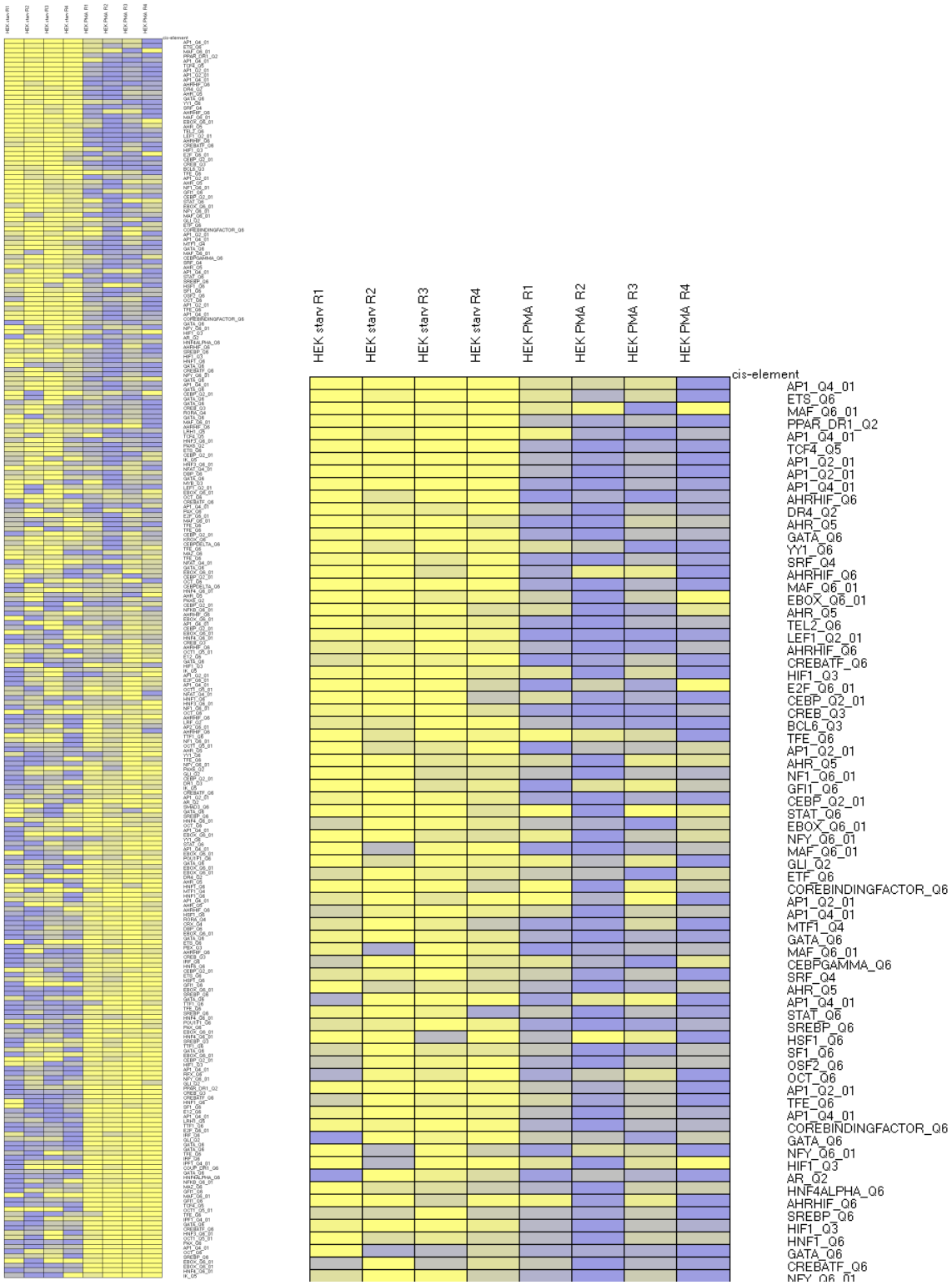


Figure3.17: heatmap with the EXT reads of the proliferation experiment with stable 293 HEK wt cells.

Transcription factors that showed a downregulation of their activity in the RNA samples with serum and PMA stimulation was e.g. GATA, E-BOX and IRF (table 3.3).

	HEK starv	HEK starv	HEK starv	HEK starv	HEK PMA	HEK PMA	HEK PMA	HEK PMA	Ratio mean
Cis-element	R1	R2	R3	R4	R1	R2	R3	R4	stim/starv
E2F_Q6_01	4430	3927	3059	3763	976	1005	806	980	0.2
IRF_Q6	6	23	17	27	2	11	5	0	0.2
GLI_Q2	275	217	200	486	84	72	74	54	0.2
GATA_Q6	8540	8187	7776	8904	1757	2354	1780	2078	0.2
GATA_Q6	471	364	334	433	115	125	73	56	0.2
IRF_Q6	2842	2571	2571	2312	593	654	527	563	0.2
IPF1_Q4_01	1047	381	332	725	122	150	136	156	0.2
GATA_Q6	94	75	77	117	14	22	21	20	0.2
HNF4ALPHA_Q6	180	154	156	160	40	45	21	31	0.2

Table 3.3: part II of the sequencing results of the TF library RNA stable 293 HEK wt proliferation experiment

3.6 TripZ-virus cell culture experiments

The measurement of the transcription factor activity using EXT based assays worked with the conventional produced stable cell lines and Next Generation Sequencing. Only one problem has turned up: human cell lines especially cancer cell lines were often not transfectable. For these cell lines a plasmid pTripZ, carrying lenti virus genes, was used. Within this plasmid a destination box for the Gateway cloning was cloned (figure 3.18).

Then the TF, Xie and Co libraries were cloned into this DEST-vector.

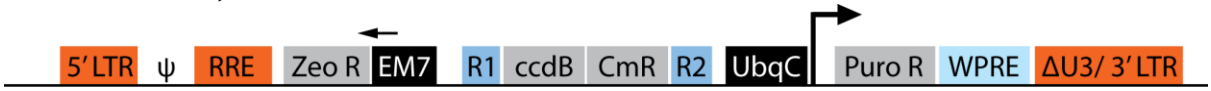


Figure 3.18: pDEST-TripZ:
Scheme of the cloned DEST-vector, the ccdB cassette was cloned in between the ubiquitin promoter and the EM7 promoter.

A cytomegalovirus (CMV) promoter driven enhanced yellow fluorescent protein (EYFP) was cloned into the pDEST-TripZ via Gateway cloning first to test the functionality of the virus (figure 3.19). All virus production experiments were done in 15 cm cell culture dishes with the cell line 293 HEK FT. The pEXPR-TripZ-EYFP was co transfected with an appropriate amount of the two packaging plasmids psPAX2 and pMD2.G as described within the methods. The cleaned up virus was titrated and 293 HEK cells were infected with a MOI of 10 (figure 3.20).

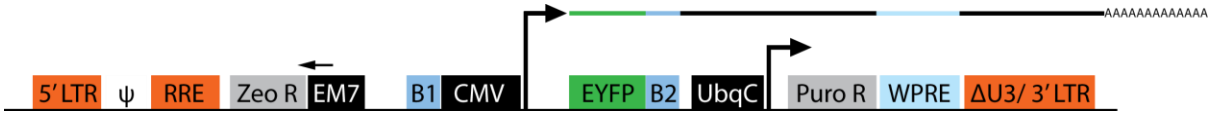


Figure 3.19: pEXPR-TripZ-EYFP
Scheme of the TripZ virus backbone carrying the EYFP gene.

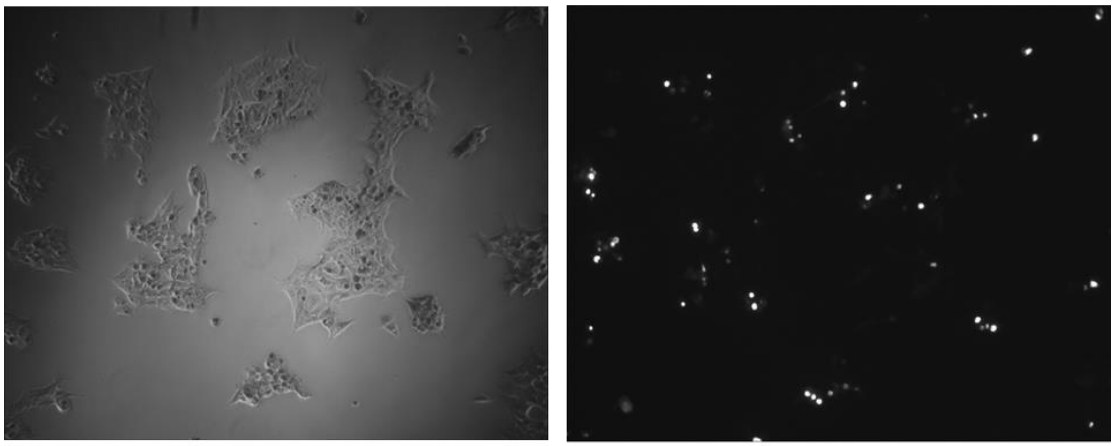


Figure 3.20: infected 293 HEK cells of the virus TripZ-EYFP

The virus libraries TF and Xie were produced after running the functionality test. These *Cis*-element EXT libraries had no firefly luciferase 3' to the EXTs, so these constructs could not be used for luciferase assays and the online luciferase measurement.

Figure 3.21: pEXPR-TripZ
Scheme of the cloned EXPR-vector containing a *Cis*-element with the coupled EXT

To control the infection of the cell lines 1 µg of the pTripZ-EYFP construct was added to 14µg of pEXPR-TripZ of the TF and Xie libraries into the TripZ virus production step. With the produced and purified virus cell lines as 293 HEK wt, HT-29 and U2OS were infected and selected with puromycine. After a short passaging process these cell lines were used for a proliferation experiment. In this experiment the cells were seeded on 10 cm cell culture dishes with a density of 2.5 to 3 million cells per dish. Then the cells were transfected with an expression vector containing the early growth response protein 1 (EGR1) promoter region coupled to the firefly luciferase - and incubated for 24 hours in normal growth media. The first samples were lysed in RLT buffer before starving, these were named t=0. The cells were starved for 18 hours with starvation media containing 0.1% FBS and the next samples were lysed after the starvation. The rest of the cells were stimulated with normal growth media containing 10% FBS and were lysed at three time points 6 hours, 12 hours and 24 hours after starvation. All samples were processed as described in the experiments before and the samples of the stable 293 HEK wt TF library were sequenced. During the experiment the firefly luciferase activity of the transfected EGR1 construct was measured. The luciferase activity showed a peak at hour 4 in the stimulation (figure 3.22).

The EXT sequencing reads were normed to the total number of reads and analyzed. In the analysis of the EXT sequencing almost 3000 different EXTs were identified in the RNA and DNA samples.

The 500 best performing EXTs were used for further analysis and heatmapping for better visualization.

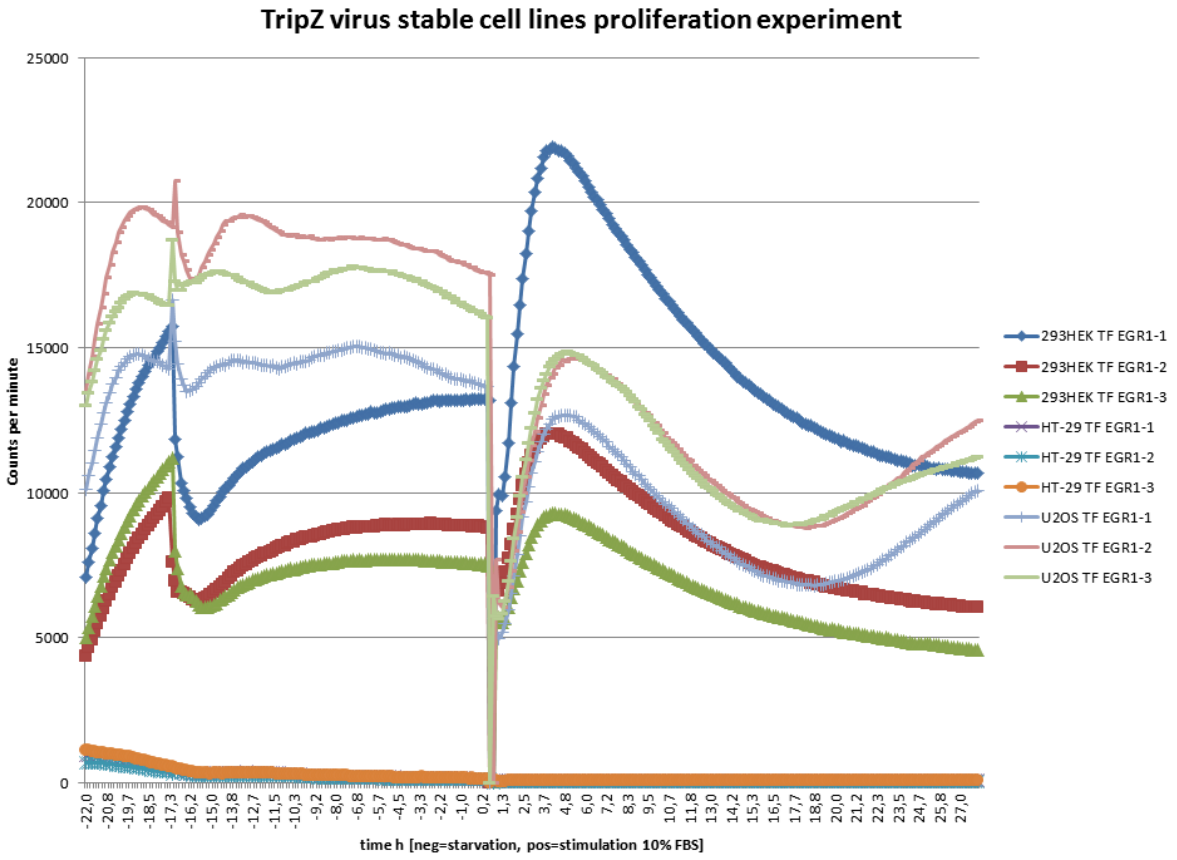


Figure 3.22: luciferase results of the TripZ virus stable cell line experiment:
The cell line HT-29 was not transfectable, so these cells had no significant luciferase reads. The cell lines 293 HEK wt and U2OS showed a peak of the EGR1-luciferase construct at hour four of the stimulation.

Comparing the RNA samples of the starvation versus the 24 h stimulation transcription factors AP1, CREB, sterol regulatory element binding protein (SREBP) and CCAAT/enhancer-binding-protein (C/EBP) were activated (figure 3.23). A downregulation was observed on transcription factors as octamer binding transcription factor (OCT) a stress sensor in cells (Tantin et al.2005), SMAD3 and cone-rod homebox protein (CRX) (figure 3.23 and table 3.4).

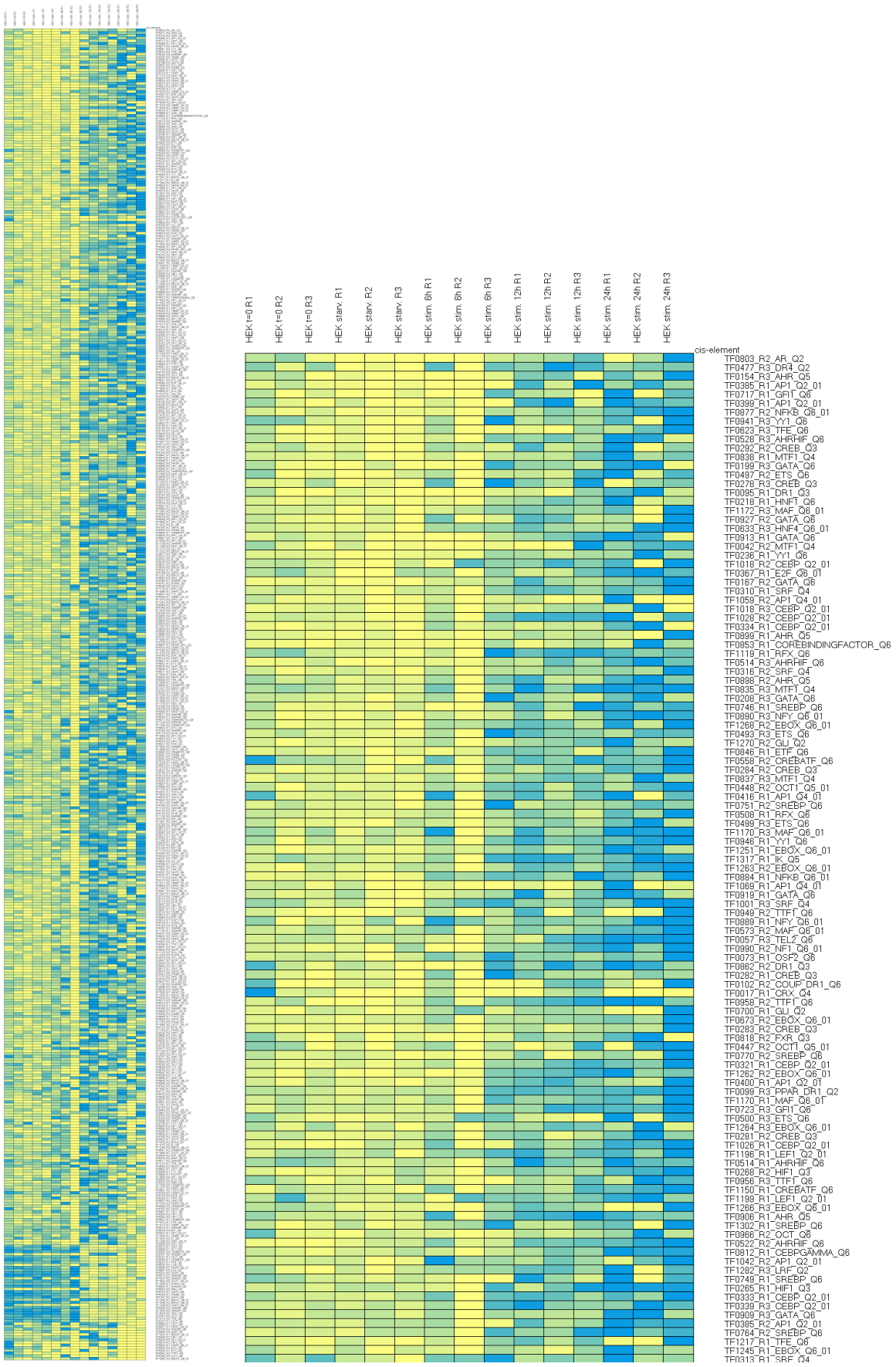


Figure 3.23: heatmap with the EXT reads of the proliferation experiment with TripZ-infected 293 HEK wt cells.

Cis-element	HEK t=0 R1	HEK t=0 R2	HEK t=0 R3	HEK starv. R1	HEK starv. R2	HEK starv. R3	HEK st. 24h R1	HEK st. 24h R2	HEK st. 24h R3	ratio 24hstim/t=0
COREBIND.F._Q6	569	448	123	726	533	412	998	7521	2339	9.525
AHRHIF_Q6	7308	4674	3579	7406	12518	12578	32756	43596	51379	8.208
MYB_Q3	646	307	175	1041	713	2196	2273	5795	1128	8.152
HIF1_Q3	1131	359	831	805	1056	2589	5390	5561	7607	7.996
EBOX_Q6_01	477	2	921	465	426	1070	2653	3477	4064	7.281
HNF3_Q6_01	478	0	89	4	741	390	1259	376	2460	7.222
GATA_Q6	570	31	1	451	917	1389	3213	1072	14	7.141
TFE_Q6	535	1	0	218	0	100	477	1215	2033	6.950
NF1_Q6_01	710	506	219	524	2	1512	3957	1809	4164	6.920
GATA_Q6	1081	624	333	515	1773	1815	2832	4849	5954	6.690
NFAT_Q4_01	475	284	205	519	951	1075	1381	1469	3405	6.489
CEBP_Q2_01	1162	896	646	648	2080	1365	5557	4754	6874	6.355
PAX6_Q2	451	105	380	1092	841	1091	964	3992	873	6.228
OCT_Q6	450	0	601	1017	745	491	1663	1069	3796	6.211
EBOX_Q6_01	611	124	380	570	345	360	3276	2538	910	6.030
SREBP_Q6	609	37	268	1147	1426	915	1601	2592	1307	6.018
CREB_Q3	1297	213	775	1124	1	1203	4257	3228	6186	5.983
E2F_Q6_01	475	1519	1981	3467	2461	3477	9065	8501	6106	5.955
DR1_Q3	923	0	1	400	203	0	2015	2908	526	5.897
NFKB_Q6_01	463	168	481	119	219	198	1668	2212	2622	5.847
AP1_Q4_01	554	462	254	33	769	1440	1642	2008	3747	5.824
CEBPGAMMA_Q6	1324	847	483	1120	988	1748	5587	4662	4881	5.701
CEBP_Q2_01	942	506	1102	601	509	2896	4633	3367	6237	5.583
CEBP_Q2_01	608	248	829	582	603	207	3488	5450	364	5.520
AR_Q2	2779	2075	2357	3280	4080	3460	10080	13115	15881	5.419
NFY_Q6_01	9132	2724	4826	3862	7325	3096	27499	32921	27387	5.264
AHRHIF_Q6	1126	252	108	382	129	195	3850	3387	536	5.231
RORA_Q4	1856	618	752	910	2328	2688	7643	2446	6281	5.074
HIF1_Q3	566	0	1	613	382	254	751	287	1836	5.069
SRF_Q4	1004	184	574	946	419	14	2131	1633	5154	5.061
SREBP_Q6	665	967	953	387	1512	1563	5699	2971	4303	5.019
HIF1_Q3	473	644	381	724	748	463	2577	2379	2379	4.897
CREB_Q3	988	66	504	661	1513	983	5752	0	1799	4.847
MTF1_Q4	792	598	396	788	1	111	4708	1434	2467	4.820
GATA_Q6	705	816	940	429	708	477	2597	3455	5808	4.819
GATA_Q6	446	78	57	18	275	0	1251	1436	79	4.761
E12_Q6	828	451	1166	647	932	1824	3736	1160	6683	4.736
AP1_Q2_01	671	204	15	339	1	0	1842	1295	1075	4.733
DR3_Q4	1967	69	493	1184	1400	1769	4831	2941	4085	4.688
TTF1_Q6	663	247	1	17	1	834	1	243	4007	4.666
HSF1_Q6	974	169	159	188	550	1568	2336	627	3105	4.661
GATA_Q6	2041	1087	2655	3143	2920	1078	11543	11988	3371	4.652
CEBP_Q2_01	2040	1101	1168	1860	896	1635	5514	4816	9705	4.650
GATA_Q6	2509	1588	2092	3629	2166	4484	5899	14703	8073	4.633

AHRHIF_Q6	1001	919	309	567	1758	580	3604	1270	5433	4.624
EBOX_Q6_01	1759	418	222	655	2046	1690	2912	7121	1051	4.620
TTF1_Q6	794	655	202	1689	5	2129	3679	199	3709	4.595
SREBP_Q6	628	5	209	646	78	245	1425	7	2433	4.590
GLI_Q2	1732	201	1545	624	669	1394	9953	3526	2472	4.586
AP1_Q2_01	865	822	1145	2116	1804	3280	5985	1438	5555	4.583
EBOX_Q6_01	499	622	435	168	682	668	2002	1282	3822	4.567
CREB_Q3	634	394	904	715	367	1697	2466	1975	4375	4.563
IPF1_Q4_01	774	685	319	506	1107	1136	1489	5262	1344	4.553
ETF_Q6	651	908	601	858	139	661	3066	1845	4904	4.544
AHRHIF_Q6	1129	168	326	335	1401	1366	3609	1953	1796	4.534
TEL2_Q6	449	741	703	308	764	655	2825	2735	2927	4.483
HNF4_Q6_01	455	351	219	264	261	31	636	1902	2053	4.479
GATA_Q6	2546	2112	3245	7232	4129	5102	14431	12591	8133	4.448
SRF_Q4	775	108	1388	1505	700	1100	3648	6294	137	4.438
PPAR_DR1_Q2	655	670	84	814	625	1020	2868	1061	2280	4.407
IK_Q5	535	2	557	962	227	489	1023	2576	1137	4.329
CREB_Q3	593	460	443	534	581	780	1276	4312	812	4.278
GATA_Q6	681	321	515	1091	199	0	3341	2388	747	4.269
HNF3_Q6_01	789	406	359	606	334	1204	4003	183	2402	4.239
SREBP_Q6	890	879	467	1083	1210	1304	2299	2346	4830	4.237
SREBP_Q6	850	210	241	223	152	513	2318	1105	2073	4.224
MTF1_Q4	606	1819	1108	1004	932	603	2796	8921	3172	4.214
GFI1_Q6	620	1654	928	1255	1105	755	3904	3630	5802	4.165
AHRHIF_Q6	1106	1220	2317	836	765	1381	5922	8270	5105	4.156
AP1_Q2_01	508	427	280	260	579	509	1946	1474	1629	4.156
AHR_Q5	836	611	531	744	1376	1060	2595	2424	3185	4.148
HNF1_Q6	509	163	429	322	83	113	3292	402	868	4.144
EBOX_Q6_01	777	324	829	902	979	1420	2860	2117	3003	4.135
ETS_Q6	794	258	345	216	26	825	1820	3130	807	4.121
NFAT_Q4_01	1586	2519	1436	3464	2200	2613	11826	3460	7464	4.106
NRSF_Q4	529	1	2	835	314	172	9	822	1351	4.102
OCT1_Q5_01	1378	23	1066	437	790	571	1489	3299	5298	4.088
NFAT_Q4_01	1406	969	1749	2401	1638	1170	3025	3654	10177	4.087
SF1_Q6	610	400	382	888	854	1541	3551	756	1366	4.075
SRF_Q4	599	904	356	372	1	680	2277	1752	3543	4.073
GATA_Q6	1693	1339	1797	2502	1538	1770	6946	3729	8887	4.051
YY1_Q6	965	0	752	303	729	275	1661	2496	2785	4.043
PAX_Q6	1032	597	144	646	482	1060	2461	3467	1223	4.033
NFKB_Q6_01	776	812	1122	831	432	803	3420	2150	5134	3.950
GFI1_Q6	811	0	451	32	15	308	3033	4	1924	3.931
E2F_Q6_01	456	794	755	940	452	1794	3355	2063	2424	3.911
NFY_Q6_01	752	37	319	143	719	3	851	1743	1705	3.880
CEBP_Q2_01	564	775	1195	1231	31	202	1849	7806	145	3.867
GATA_Q6	473	1015	500	737	129	103	4159	3250	279	3.867

Table 3.4: extraxt of the EXT reads of the proliferation experiment with TripZ infected 293 HEK wt cells.

4. Discussion

4.1 Reporter gene libraries

The *Cis*-element libraries TF, Xie and Co coupled to unique EXTs were created for the measurement of transcription factor activity in parallel in living cells. The sensitivity and specificity of the EXTs as well as the dynamic response and stable readout in cell culture experiments was previously tested and shown by Anna Botvinnik (Botvinnik et al. 2010).

The construct size of the oligonucleotide was 180 base pairs. The length and quality of the oligonucleotide was limited by the technique - the on-chip DNA synthesis - itself.

The TF library was designed for the measurement of transcription factor activity in human or murine cell lines. The Xie library enables the discovery of unknown transcription factor binding sites. With the Co library readouts of cooperating and interacting transcription factors and signalling pathways might be possible.

4.2 DNA quality

In a first analysis the complexity and functionality of random single clones were evaluated via sequencing.

In the sequencing of the random selected single clones of all three libraries 63% (78 clones) of all 122 single clones were free of mutation, 16 % (20 clones) had one mutation. These results led to the conclusion that approximately 70% out of these single clones were functional. During the analysis of the deep sequencing experiment we saw that 10% of all constructs were coupled to a wrong EXT and out of the Xie library only 53% of all EXTs could be mapped. The on-chip inkjet printing method could have led to the mismatching of *Cis*-elements with EXTs and the loss of almost 50% out of the Xie library. A further reason for the construct loss of the Xie library might be the concentration of repetitive sequences. There either the in the printing on-chip method or during the PCR amplification steps could have led to a significant quality loss.

4.3 Proof-of-principle

The dynamic response of extrinsic transcription factor activity was tested in a series of proof-of-principle experiments. In co-transfections experiments selected *Cis*-elements were used for luciferase readout with their corresponding transcription factor. These data were normed by co-transfected renilla luciferase.

The specificity of the *Cis*-elements was random negative tested with the HIF readout (figure 3.8). The co-transfected transcription factors are not able to bind on the HIF *Cis*-element. The changes in the relative Luciferase Units (rLUs) were not significant. Respective to the specificity the HIF experiment should have been repeated with the correct binding partners – the transcription factors HIF1a and ARNT in a growth medium where $CoCl_2$ is added.

The dynamic response of the luciferase assays was tested in a luciferase assay with the *Cis*-element NfkB using different amounts of DNA of the co-transfected transcription RelA (figure 3.6). Taken together, the functionality of the constructs could be validated in luciferase assay readout as a basis for the comprehensive follow-up studies.

4.4 Stable cell culture and TripZ-virus cell culture experiments

Plasmids containing *Cis*-element libraries coupled to unique EXTs were transfected to human cell lines 293 HEK wt, MCF-7, WiDr and SH-SY5Y. These cell lines were cultured under different culture selection conditions, respectively. Genomic DNA and RNA was isolated, amplified and analyzed with Next Generation Sequencing to (i) detect the DNA input and (ii) detect reporter expression differences as a change of the intrinsic transcription factor activity under different growth conditions. Out of the stable 293 HEK and SH-SY5Y cells 155 different EXTs in the SH-SY5Y and 266 different EXTs in the 293HEK wt cells could be identified after normalization. Some of these EXTs show robust DNA and RNA signals in every biological replicate (table 3.1; 3.2 and table 3.3). However the EXTs of the technical replicates of the *Cis*-element libraries could be identified only in little cases with signal intensity above background level (data not shown). This finding could be due to the problem of the efficiency of the generation of stable cell lines and the aspect of the localization of the integration (Weis et al. 1991).

In the virus cell culture experiment 293HEK wt, U2OS and HT-29 were infected via TripZ lentivirus containing the *Cis*-element libraries. A proliferation assay was performed in different growth media, respectively. The further analysis was done as described above. Via Next Generation Sequencing 2938 different EXTs of the TF library could be detected. The 500 best performing EXTs were analyzed (table 3.4). In summary, the usage of virus systems increases dramatically the efficiency of the stable integration especially in non-

transfectable cell lines, however the transcription factor activity pattern of the conventional stable cells in the 293HEK wt cell line could not have been validated (table 3.2 and table 3.4).

4.5 Selection of potential *Cis*-element candidates

Cis-element reporter gene assays are a well understood and an often used method for direct monitoring of transcription factor activity. Transcription factors have the ability to bind to specific DNA regions and work as activators or repressors. Here we can only focus on transcription factors that act as activators. All transcription factors, binding sites and functions are still not known, so there is a demand on high-quality multiplexing detection assays. With the described *Cis*-element assay based on EXTs extrinsic transcription factor activity could be precisely monitored and validated in several luciferase assays. In the cell culture experiments some transcription factors could represent cell line specific prediction markers, respectively.

In the SH-SY5Y cell line experiment where the cells were stimulated with serum and PMA after a starvation period transcription factors as IK and CRX were upregulated. In contrast to the 293HEK wt cell line samples with the same treatment these transcription factors were downregulated. In the 293HEK wt virus cell line this effect was not validated. This might be due to the missing PMA stimulation; all virus generated cell lines were serum only stimulated after a comparable starvation period. Ikaros (IK), a member of the zinc finger transcription factors that also plays an important role in hematological malignancies (John and Ward 2011; Rebollo and Schmitt 2003). In this case follow-up experiments with individual and complementary assays for the validation of these findings are required.

In SH-SY5Y and 293HEK wt cells multiple *Cis*-elements of the transcription factor AP1 showed increased signal intensities. AP1 is a transcription factor that is active at multiple situations in the cell, as well as during proliferation, transformation and cell death (Karin et al. 1997; Shaulian and Karin 2002). follow-up experiments are essential to validate this observation.
NFAT, an important member in immune response and in calcium signaling (Rao et al. 1997), showed in the SH-SY5Y cells a downregulation. In the 293HEK wt cells GATA, a member of the zinc finger transcription factor that is important for cell maturation, cell proliferation arrest and cell survival (Zheng and Blobel 2010) showed a massively downregulation.
The virus cell culture experiments were done with 293HEK wt, U2OS and HT-29. A serum starvation and serum stimulation as a proliferation assay was done. However only the cell samples of the 293HEK wt cell line was analyzed by Next Generation Sequencing, so there it is impossible to make comparisons of the signal patterns between the different cell lines. Taken together, this approach of multiplexing *Cis*-element libraries coupled to unique EXTs could show some effects on transcription factor activity in cell proliferation that are described in literature. Although a set of first proof-of-principle experiments could be performed successfully within this thesis, further validation and optimization is essential to improve these *Cis*-element based EXTassays to increase robustness for future applications.

5. Conclusion

In the present study an upscaling of a described method called EXTassays was established. In three different *Cis*-element libraries all currently described and validated as well as phylogenetic conserved potential transcription factor binding sites were covered for a simultaneous monitoring of the activity of transcription factors in living cells. Due to the replacement of classical reporter gene assays by unique DNA barcodes so called EXT assays or EXTassays a multiplexed quantification of intrinsic transcription factor activity was performed with the help of Next Generation Sequencing.

Using classical generated and virus generated stable cell lines in cell culture experiments, proliferation assays including serum starvation and cell stimulation via high serum concentrations were performed with the aim to generate transcription factor activity patterns of different cancer cell lines under different situations, chemotherapy treatment e.g. to find new candidates of transcription factors as predictable markers for cell surviving mechanisms or cell proliferation e.g. several changes in the cell specific expression patterns of the EXTs were found depending on the experiment conditions. However, follow-up experiments are pending to validate these initial findings.

These data suggest that virus systems could be the more reliable method for generation of stable cell lines with highly complex reporter gene libraries. Comparing the expression patterns of the virus generated cell lines is not possible yet, these samples have to be analyzed with Next Generation Signaling.

In summary, the applicability of an EXT-based *Cis*-element reporter assay system to quantitatively measure the effects of cellular events on transcription factor activity could be demonstrated. However, as these experiments represent a first proof-of-principle experiment for the simultaneous assessment of possibly all cellular transcription factor activities further optimization is required to increase the robustness of this promising novel assay system.

6.0 Supplement
6.1 *Cis*-element libraries

TF: (abstract)

unique ID	matrix_name	TF sequence	species
TF0001	V$VDR_Q3	GAGTTCACCGGGTGTGAGTTCACCGGGTGTGAGTTC	Homo sapiens
TF0002	V$VDR_Q3	GAGTCAGCGAGGTGAGAGTCAGCGAGGTGAGAGTCA	Homo sapiens
TF0003	V$VDR_Q3	GGGTGAACGGGGGCAGGGTGAACGGGGGCAGGGTGA	Homo sapiens
TF0004	V$VDR_Q3	AGGGAGATTGGTTCAAGGGAGATTGGTTCAAGGGAG	Homo sapiens
TF0005	V$CEBPDELTA_Q6	GTTTGCGCCACTGTTTGCGCCACTGTTTGCGCCACT	Homo sapiens
TF0006	V$CEBPDELTA_Q6	CATTTCGTAATTCATTTCGTAATTCATTTCGTAATT	Homo sapiens
TF0007	V$CEBPDELTA_Q6	CATTGCACAATCCATTGCACAATCCATTGCACAATC	Homo sapiens
TF0008	V$CEBPDELTA_Q6	GATTACATCACTGATTACATCACTGATTACATCACT	Homo sapiens
TF0009	V$CEBPDELTA_Q6	AATGACATCACAAATGACATCACAAATGACATCACA	Homo sapiens
TF0010	V$CEBPDELTA_Q6	AATTGCGTAAGCAATTGCGTAAGCAATTGCGTAAGC	Homo sapiens
TF0011	V$CEBPDELTA_Q6	CATTTCGTCACACATTTCGTCACACATTTCGTCACA	Homo sapiens
TF0012	V$CRX_Q4	TCCATAAGACGATTCCATAAGACGATTCCATAAGAC	Homo sapiens
TF0013	V$CRX_Q4	CCTGTAATCCCAGCCTGTAATCCCAGCCTGTAATCC	Homo sapiens
TF0014	V$CRX_Q4	TGCTTAATGCTTATGCTTAATGCTTATGCTTAATGC	Homo sapiens
TF0015	V$CRX_Q4	TCATTTAGCTGTTTCATTTAGCTGTTTCATTTAGCT	Homo sapiens
TF0016	V$CRX_Q4	TGATTAAACTCAGTGATTAAACTCAGTGATTAAACT	Homo sapiens
TF0017	V$CRX_Q4	CAGCTAATGCGACCAGCTAATGCGACCAGCTAATGC	Homo sapiens
TF0018	V$CRX_Q4	TGCATAACACCAGTGCATAACACCAGTGCATAACAC	Homo sapiens
TF0019	V$CRX_Q4	CTGTTGATCCCCTCTGTTGATCCCCTCTGTTGATCC	Homo sapiens
TF0020	V$CRX_Q4	GAACTAATCCCCTGAACTAATCCCCTGAACTAATCC	Homo sapiens
TF0021	V$CRX_Q4	TTCCTAATCTCACTTCCTAATCTCACTTCCTAATCT	Homo sapiens
TF0022	V$DBP_Q6	AGCACACAGCACACAGCACACAGCACACAGCACACA	Homo sapiens
TF0023	V$DBP_Q6	AGCAAACAGCAAACAGCAAACAGCAAACAGCAAACA	Homo sapiens
TF0024	V$DBP_Q6	AGCAAACAGCAAACAGCAAACAGCAAACAGCAAACA	Homo sapiens
TF0025	V$DBP_Q6	AGCAAAAAGCAAAAAGCAAAAAGCAAAAAGCAAAAA	Homo sapiens
TF0026	V$FXR_Q3	CAGGGTGAATAACCCAGGGTGAATAACCCAGGGTGA	Homo sapiens
TF0027	V$FXR_Q3	CAAGGTCATTAACTCAAGGTCATTAACTCAAGGTCA	Homo sapiens
TF0028	V$HNF4ALPHA_Q6	GTGGACTTAGCCCGTGGACTTAGCCCGTGGACTTAG	Homo sapiens
TF0029	V$HNF4ALPHA_Q6	GCAATCTTTGACCGCAATCTTTGACCGCAATCTTTG	Homo sapiens
TF0030	V$HNF4ALPHA_Q6	GTGACCTTTGCCCGTGACCTTTGCCCGTGACCTTTG	Homo sapiens
TF0031	V$HNF4ALPHA_Q6	CAGAGCTTTGTCTCAGAGCTTTGTCTCAGAGCTTTG	Homo sapiens
TF0032	V$HNF4ALPHA_Q6	CAGAACCTTTAAGCAGAACCTTTAAGCAGAACCTTT	Homo sapiens
TF0033	V$LXR_Q3	TGGGGTCATTGTCGGGCATGGGGTCATTGTCGGGCA	Homo sapiens
TF0034	V$LXR_Q3	CAGGGTCACTGGCGGTCACAGGGTCACTGGCGGTCA	Homo sapiens
TF0035	V$MAZ_Q6	GGGGAGGGGGGGAGGGGGGAGGGGGGAGGGGGGG	Homo sapiens
TF0036	V$MAZ_Q6	GGGGAGGGGGGGAGGGGGGAGGGGGGAGGGGGGG	Homo sapiens
TF0037	V$MAZ_Q6	GGGGCGAGGGGGCGAGGGGCGAGGGGCGAGGGGG	Homo sapiens
TF0038	V$MAZ_Q6	AGGGAGCGAGGGAGCGAGGGAGCGAGGGAGCGAGGG	Homo sapiens
TF0039	V$MAZ_Q6	GGGGAGGGGGGGAGGGGGGAGGGGGGAGGGGGGG	Homo sapiens
TF0040	V$MAZ_Q6	TGTGAGGGTGTGAGGGTGTGAGGGTGTGAGGGTGTG	Homo sapiens
TF0041	V$MAZ_Q6	AGGGAGGGAGGGAGGGAGGGAGGGAGGGAGGGAGGG	Homo sapiens
TF0042	V$MTF1_Q4	TTTGCACTCGTCCCTTTGCACTCGTCCCTTTGCACT	Homo sapiens
TF0043	V$MTF1_Q4	CCTGCACACGCCCCCCTGCACACGCCCCCCTGCACA	Homo sapiens
TF0044	V$MTF1_Q4	TCTGCACACGGGCCTCTGCACACGGGCCTCTGCACA	Homo sapiens
TF0045	V$MTF1_Q4	TGTGCACACGGCGGTGTGCACACGGCGGTGTGCACA	Homo sapiens
TF0046	V$MTF1_Q4	GGTGCGCCCGGCCCGGTGCGCCCGGCCCGGTGCGCC	Homo sapiens
TF0047	V$MTF1_Q4	GCTGCACCCAGCCCGCTGCACCCAGCCCGCTGCACC	Homo sapiens
TF0048	V$SP3_Q3	ACCCTGGGGGCGGGACCCTGGGGGCGGGACCCTGGG	Homo sapiens
TF0049	V$SP3_Q3	ATCCCTGGGAGGGGATCCCTGGGAGGGGATCCCTGG	Homo sapiens
TF0050	V$TCF4_Q5	GCTTTGATGCTTTGATGCTTTGATGCTTTGATGCTT	Homo sapiens
TF0051	V$TCF4_Q5	CCTTTGATCCTTTGATCCTTTGATCCTTTGATCCTT	Homo sapiens
TF0052	V$TCF4_Q5	CCCTTTAGCCCTTTAGCCCTTTAGCCCTTTAGCCCT	Homo sapiens
TF0053	V$TCF4_Q5	CCTTTGAACCTTTGAACCTTTGAACCTTTGAACCTT	Homo sapiens
TF0054	V$TCF4_Q5	ACTTTGAAACTTTGAAACTTTGAAACTTTGAAACTT	Homo sapiens
TF0055	V$TCF4_Q5	GCTTTGAAGCTTTGAAGCTTTGAAGCTTTGAAGCTT	Homo sapiens
TF0056	V$TEL2_Q6	CTACTTCCTGCTACTTCCTGCTACTTCCTGCTACTT	Homo sapiens
TF0057	V$TEL2_Q6	TCACTTCCTGTCACTTCCTGTCACTTCCTGTCACTT	Homo sapiens
TF0058	V$E12_Q6	CGCAGATGTCCCGCAGATGTCCCGCAGATGTCCCGC	Homo sapiens
TF0059	V$ETF_Q6	GAGGAGGGAGGAGGGAGGAGGGAGGAGGGAGGAGGG	Homo sapiens
TF0060	V$ETF_Q6	GCGGCCGGCGGCCGGCGGCCGGCGGCCGGCGGCCGG	Homo sapiens
TF0061	V$ETF_Q6	GCGGAGGGCGGAGGGCGGAGGGCGGAGGGCGGAGGG	Homo sapiens
TF0062	V$SMAD3_Q6	TGTCTGACTTGTCTGACTTGTCTGACTTGTCTGACT	Homo sapiens
TF0063	V$SMAD3_Q6	TGTCTGTCTTGTCTGTCTTGTCTGTCTTGTCTGTCT	Homo sapiens
TF0064	V$SMAD3_Q6	CATCTGTCTCATCTGTCTCATCTGTCTCATCTGTCT	Homo sapiens
TF0065	V$SMAD3_Q6	CGGCTGACTCGGCTGACTCGGCTGACTCGGCTGACT	Homo sapiens
TF0066	V$SMAD3_Q6	TGTCTGTCTTGTCTGTCTTGTCTGTCTTGTCTGTCT	Homo sapiens
TF0067	V$USF2_Q6	CAGGCGCAGGCGCAGGCGCAGGCGCAGGCGCAGGCG	Homo sapiens
TF0068	V$SF1_Q6	GGACCTTGGGACCTTGGGACCTTGGGACCTTGGGAC	Homo sapiens
TF0069	V$SF1_Q6	TTACCTTGTTACCTTGTTACCTTGTTACCTTGTTAC	Homo sapiens
TF0070	V$SF1_Q6	TAACCTTGTAACCTTGTAACCTTGTAACCTTGTAAC	Homo sapiens
TF0071	V$OSF2_Q6	ACCACATGACCACATGACCACATGACCACATGACCA	Homo sapiens
TF0072	V$OSF2_Q6	ACCGCAAAACCGCAAAACCGCAAAACCGCAAAACCG	Homo sapiens

Xie: (abstract)

Unique ID	Xie-name	Xie sequence
X001	V$NRF1_Q6	GCGCATGCGTATAGCGCATGCGTATAGCGCATGCGTATAGCGCATGCGTA
X002	-	AAGCATGCCTATAAAGCATGCCTATAAAGCATGCCTATAAAGCATGCCTA
X003	V$NRF1_Q6	CGCGTGTGCATTATACGCGTGTGCATTATACGCGTGTGCATTATACGCGT
X004	-	GCGCATACGCAGATAGCGCATACGCAGATAGCGCATACGCAGATAGCGCA
X005	-	GCGCACGTCCAGCATAGCGCACGTCCAGCATAGCGCACGTCCAGCATAGC
X006	-	CAGCATGTATACAGCATGTATACAGCATGTATACAGCATGTATACAGCAT
X007	-	ACGCAGCCGAAGTATAACGCAGCCGAAGTATAACGCAGCCGAAGTATAAC
X008	-	CATGTCCAGCATACATGTCCAGCATACATGTCCAGCATACATGTCCAGCA
X009	-	GCGCAAGCTCATAGCGCAAGCTCATAGCGCAAGCTCATAGCGCAAGCTCA
X010	-	AGCATGTAATAAGCATGTAATAAGCATGTAATAAGCATGTAATAAGCATG
X011	V$MYC_Q2	CACGTGATACACGTGATACACGTGATACACGTGATACACGTGATACACGT
X012	V$USF_Q6_01	TCACGTGATATCACGTGATATCACGTGATATCACGTGATATCACGTGATA
X013	-	GCCACGTCATAGCCACGTCATAGCCACGTCATAGCCACGTCATAGCCACG
X014	V$MYCMAX_B	GCCACGCCATAGCCACGCCATAGCCACGCCATAGCCACGCCATAGCCACG
X015	V$USF_C	AGCACGTGATAAGCACGTGATAAGCACGTGATAAGCACGTGATAAGCACG
X016	-	ACGCATGTGATAACGCATGTGATAACGCATGTGATAACGCATGTGATAAC
X017	-	GTTTCCTGTGATAGTTTCCTGTGATAGTTTCCTGTGATAGTTTCCTGTGA
X018	V$ELK1_02	CCGGAAGCATACCGGAAGCATACCGGAAGCATACCGGAAGCATACCGGAA
X019	-	CCCGGAAAATACCCGGAAAATACCCGGAAAATACCCGGAAAATACCCGGA
X020	-	CATCCGGGTCATACATCCGGGTCATACATCCGGGTCATACATCCGGGTCA
X021	-	GACCTGGAAAATAGACCTGGAAAATAGACCTGGAAAATAGACCTGGAAAA
X022	V$ELK1_02	CCGGAAGCATACCGGAAGCATACCGGAAGCATACCGGAAGCATACCGGAA
X023	-	ACATACGGATAACATACGGATAACATACGGATAACATACGGATAACATAC
X024	V$ELK1_02	GGTTCCGGATAGGTTCCGGATAGGTTCCGGATAGGTTCCGGATAGGTTCC
X025	-	AAGTCCCGGATAAAGTCCCGGATAAAGTCCCGGATAAAGTCCCGGATAAA
X026	V$PEA3_Q6	AACATCCGATAAACATCCGATAAACATCCGATAAACATCCGATAAACATC
X027	-	TTCCATATGTATATTCCATATGTATATTCCATATGTATATTCCATATGTA
X028	V$RP58_01	CAGATGTTATACAGATGTTATACAGATGTTATACAGATGTTATACAGATG
X029	-	CTGGTTCCGATACTGGTTCCGATACTGGTTCCGATACTGGTTCCGATACT
X030	-	ACTATAGGGCCCAATAACTATAGGGCCCAATAACTATAGGGCCCAATAAC
X031	-	ACTACGTTACCCATAACTACGTTACCCATAACTACGTTACCCATAACTAC
X032	-	ACTACCTTTCCCAATAACTACCTTTCCCAATAACTACCTTTCCCAATAAC
X033	-	AGACTACAATAAGACTACAATAAGACTACAATAAGACTACAATAAGACTA
X034	-	AGACTCCATGTATAAGACTCCATGTATAAGACTCCATGTATAAGACTCCA
X035	-	GACGCCATTATAGACGCCATTATAGACGCCATTATAGACGCCATTATAGA
X036	-	AAACCACGCTTCCCATAAAACCACGCTTCCCATAAAACCACGCTTCCCAT
X037	-	GGATTACATAGGATTACATAGGATTACATAGGATTACATAGGATTACATA
X038	V$NFY_Q6_01	GATTGGCATAGATTGGCATAGATTGGCATAGATTGGCATAGATTGGCATA
X039	V$NFY_01	GGCCAATCAATAGGCCAATCAATAGGCCAATCAATAGGCCAATCAATAGG
X040	V$NFY_01	GCCCAATGAATAGCCCAATGAATAGCCCAATGAATAGCCCAATGAATAGC
X041	V$PBX1_02	TGATTGACATATGATTGACATATGATTGACATATGATTGACATATGATTG
X042	V$NFY_Q6_01	TAATTGGTATATAATTGGTATATAATTGGTATATAATTGGTATATAATTG
X043	-	TCCAATGAATATCCAATGAATATCCAATGAATATCCAATGAATATCCAAT
X044	-	GCCCAATCAATAGCCCAATCAATAGCCCAATCAATAGCCCAATCAATAGC
X045	V$CDP_01	CCAATAGATACCAATAGATACCAATAGATACCAATAGATACCAATAGATA
X046	-	TAATTGGATATATAATTGGATATATAATTGGATATATAATTGGATATATA
X047	-	CGACCAATATACGACCAATATACGACCAATATACGACCAATATACGACCA
X048	V$SP1_Q6	GGGCGGGATAGGGCGGGATAGGGCGGGATAGGGCGGGATAGGGCGGGATA
X049	V$SP1_Q6	GGGCGGATAGGGCGGATAGGGCGGATAGGGCGGATAGGGCGGATAGGGCG
X050	V$SP1_Q6	AGGCGGGGCATAAGGCGGGGCATAAGGCGGGGCATAAGGCGGGGCATAAG
X051	V$SP1_Q4_01	GGGAGGGATAGGGAGGGATAGGGAGGGATAGGGAGGGATAGGGAGGGATA
X052	-	GGGCGGAGTATAGGGCGGAGTATAGGGCGGAGTATAGGGCGGAGTATAGG
X053	V$SP1_Q6	GGGCGGGATATAGGGCGGGATATAGGGCGGGATATAGGGCGGGATATAGG
X054	-	CAGGGGGCGCATACAGGGGGCGCATACAGGGGGCGCATACAGGGGGCGCA
X055	-	AGGTGGGGCATAAGGTGGGGCATAAGGTGGGGCATAAGGTGGGGCATAAG
X056	-	GCCCCTCCCATAGCCCCTCCCATAGCCCCTCCCATAGCCCCTCCCATAGC
X057	-	AGGCGTCGCTGATAAGGCGTCGCTGATAAGGCGTCGCTGATAAGGCGTCG
X058	V$EGR_Q6	GGGGGCGATAGGGGGCGATAGGGGGCGATAGGGGGCGATAGGGGGCGATA
X059	V$AP1_C	TGACTCAATATGACTCAATATGACTCAATATGACTCAATATGACTCAATA
X060	V$BACH2_01	TGAGTCAATATGAGTCAATATGAGTCAATATGAGTCAATATGAGTCAATA
X061	-	TGCGTCAATATGCGTCAATATGCGTCAATATGCGTCAATATGCGTCAATA
X062	V$BACH2_01	TGACTCACATATGACTCACATATGACTCACATATGACTCACATATGACTC
X063	-	TGACTCATCATATGACTCATCATATGACTCATCATATGACTCATCATATG
X064	-	TGTCAGTCAATATGTCAGTCAATATGTCAGTCAATATGTCAGTCAATATG
X065	-	TGATCCAGAATATGATCCAGAATATGATCCAGAATATGATCCAGAATATG
X066	-	TGAATAACAGATATGAATAACAGATATGAATAACAGATATGAATAACAGA
X067	-	AATTAATCAATAAATTAATCAATAAATTAATCAATAAATTAATCAATAAA
X068	-	TGACTCACCTATATGACTCACCTATATGACTCACCTATATGACTCACCTA
X069	-	TGCAGCAATATGCAGCAATATGCAGCAATATGCAGCAATATGCAGCAATA
X070	-	TGATACAAATATGATACAAATATGATACAAATATGATACAAATATGATAC
X071	-	TCTCGCGACAATATCTCGCGACAATATCTCGCGACAATATCTCGCGACAA
X072	-	ATCGCGAGAATAATCGCGAGAATAATCGCGAGAATAATCGCGAGAATAAT
X073	V$ATF3_Q6	TGATGTCAATATGATGTCAATATGATGTCAATATGATGTCAATATGATGT
X074	-	GTGAATCCACATAGTGAATCCACATAGTGAATCCACATAGTGAATCCACA
X075	-	TGATATAATCATATGATATAATCATATGATATAATCATATGATATAATCA
X076	-	TTTCATCAATATTTCATCAATATTTCATCAATATTTCATCAATATTTCAT
X077	-	GCTGATTTCAATAGCTGATTTCAATAGCTGATTTCAATAGCTGATTTCAA
X078	-	TGACGTGACAATATGACGTGACAATATGACGTGACAATATGACGTGACAA

Co: (abstract)

unique ID	Co_name	Co sequence	species
Co001		GCAAACATGTCTGGACCTCTAGACAATAGCAAACAT	Homo sapiens
Co002	MyoD/Sp1	GCCCCCACCCCTGCCCCATACAACTGACATAGCCCC	Homo sapiens
Co003		GACCAAATAAGGATACCCCACCCCTGCCATACCAAC	Homo sapiens
Co004	HNF4/HNF1	CTTAGCCCCTGTTTATAGGGTGACCTTGGTTAATAT	Homo sapiens
Co005	HNF4/HNF3	CTGAACCCTTGACCCCTGCCCTATATGCCCACTCTA	Homo sapiens
Co006	Ebox/HNF4	AGGTGATCAAATGACCAGGTGATATTCAACCTTTAC	Homo sapiens
Co007	Ebox/T3R	AGTCCTGTCACCTGATAAGTCCTGTCACCTGATAAG	Homo sapiens
Co008	CEBP/HNF4	AGGCGCCCTTTGGACCTTTTGCAATCCTGGATAAGG	Homo sapiens
Co009		TGACCTTTGCCCAGATATGACCTTTGCCCAGATATG	Homo sapiens
Co010	HNF1/IL6REBP	CTGGGAAATAATTAAAATACTGGGAAATAATTAAAA	Homo sapiens
Co011	CEBP/HNF1	GTTGCTTAAATAATTAAATATTAACATAGTTGCTTA	Homo sapiens
Co012	CEBP/IL6REBP	CTGGGAAATAGTTGCTTAAATACTGGGAAATAGTTG	Homo sapiens
Co013		TGTCCATATTAGGATATGTCCATATTAGGATATGTC	Homo sapiens
Co014	Ets/SRF	CAGGATGATACCATATTAGGATACAGGATGATACCA	Homo sapiens
Co015		CAGGATGATACCATATTAGGACATATACAGGATGAT	Homo sapiens
Co016		AGTTCCCGTCAATATACAGGATGATACCATATTAGG	Homo sapiens
Co017	SRF/HMG-I(Y)	CCATATTAGGATACCATATTAGGATACCATATTAGG	Homo sapiens
Co018	AP1/Smad	TGACATCAATACAGACAGACAGAATATGACATCAAT	Homo sapiens
Co019	AP1/Ets	GAGGATGTATATGAGTCAATAGAGGATGTATATGAG	Homo sapiens
Co020	AP1/Smad	TGAGTCAGACATATGAGTCAGACATATGAGTCAGAC	Homo sapiens
Co021	HNF1/HNF1	TTGTAATAAATAACTCAATATGGAAAATTATTTACA	Homo sapiens
Co022	CEBP/NFkappaB	GTGGCGCAAACTCCCTTATAGTGGCGCAAACTCCCT	Homo sapiens
Co023	Ebox/Ets	GGAAATACAGCTGATAGGAAATACAGCTGATAGGAA	Homo sapiens
Co024	NFkappaB/HMGIY	GGATATTCCCATAGGATATTCCCATAGGATATTCCC	Homo sapiens
Co025	NFkappaB/HMGIY	GGGAAAGTTTTATAGGGAAAGTTTTATAGGGAAAGT	Homo sapiens
Co026	NFkappaB/HMGIY	GGGGATTTCCTATAGGGGATTTCCTATAGGGGATTT	Homo sapiens
Co027	NFkappaB/NFkappaB	GGGAAAGTTTTATAGGGGATTTCCTATAGGGAAAGT	Homo sapiens
Co028	AP1/HMGIY	CTGACATCAATATTTTAATACTGACATCAATATTTT	Homo sapiens
Co029	AP1/NFkappaB	CTGACATCAATAGGGGATTTCCTATACTGACATCAA	Homo sapiens
Co030	AP1/NFkappaB	CTGACATCAATAGGATATTCCCATACTGACATCAAT	Homo sapiens
Co031	AP1/NFkappaB	CTGACATCAATAGGGAAAGTTTTATACTGACATCAA	Homo sapiens
Co032		TGGATATTCCCGGGAAAGTTTTTATATGGATATTCC	Homo sapiens
Co033		CCCCTGCTCTGACCCCGGGTGGCCCCTACCCCTGGC	Homo sapiens
Co034		TCTGACCTCTCGACCCTACCGGGCCTGAGGCCACAT	Homo sapiens
Co035	AhR/HNF4	TACGTGCTATATCTGACCTCTCGACCTATATACGTG	Homo sapiens
Co036	AP1/GATA	TTATCTATATGACTAAATATTATCTATATGACTAAA	Homo sapiens
Co037		TTATCTATAGCACGTATAGTGACTAAATATTATCTA	Homo sapiens
Co038	CEBP/NFkappaB	CAGAGATTCCATAATTTCACAAAATACAGAGATTCC	Homo sapiens
Co039	GATA/TSEB	CCTAAGGGATAAGATAAATACCTAAGGGATAAGATA	Homo sapiens
Co040	CREB/GATA	AGATAAATATGACGTCAATAAGATAAATATGACGTC	Homo sapiens
Co041	AP1/NFAT	GGAGCCCCTGAGTCAATAGGAGCCCCTGAGTCAATA	Homo sapiens
Co042	AP1/NFAT	TGATGTCATCTTTCCAATATGATGTCATCTTTCCAA	Homo sapiens
Co043	AP1/NFAT	TGACTCTATACTTTCCTATATGACTCTATACTTTCC	Homo sapiens
Co044	AP1/NFAT	TTAATCATTTCCTCATATTAATCATTTCCTCATATT	Homo sapiens
Co045		AGTTCCCCATAATTAATCATTTCCTCATAAGTTCCC	Homo sapiens
Co046	NFkappaB/HMGIY	GGAGATTCCAATAGGAGATTCCAATAGGAGATTCCA	Homo sapiens
Co047		GGAGATTCCAATAGGAGATTCCAATAGGAGATTCCA	Homo sapiens
Co048		GCTTTCCTATAGCTTTCCTATAGCTTTCCTATAGCT	Homo sapiens
Co049	PU1/IRF	GTTTTCATTTCCTCATAGTTTTCATTTCCTCATAGT	Homo sapiens
Co050	IRF/STAT	TTCTGATAAATAAGAAAAGGAAACCATATTCTGATA	Homo sapiens
Co051	NFkappaB/HMGIY	GGGAATTTCCATAGGGAATTTCCATAGGGAATTTCC	Homo sapiens
Co052		TGGGAGGAGCATATTATCCATATATGGGAGGAGCAT	Homo sapiens
Co053	Pit/Pit	TTATCCATATAATGCATAAATATTATCCATATAATG	Homo sapiens
Co054	AP2/NF1	TGGCCTGCGGCCAGAATATGGCCTGCGGCCAGAATA	Homo sapiens
Co055	CEBP/Stat	CTGGAAAATACTGGAAATATACTGGAAAATACTGGA	Homo sapiens
Co056	NFkappaB/HMGIY	GGGAAATTCCATAGGGAAATTCCATAGGGAAATTCC	Homo sapiens
Co057	IRF/NFkappaB	GAGAAGTGAAAGTGGGAAATTCCATAGAGAAGTGAA	Homo sapiens
Co058	AP1/HMGIY	TGACATAGGAAAAATATGACATAGGAAAAATATGAC	Homo sapiens
Co059	AP1/HMGIY	TAAATGACATAGATATAAATGACATAGATATAAATG	Homo sapiens
Co060	AP1/NFkappaB	TGACATAGATAGGGAAATTCCTCATATGACATAGAT	Homo sapiens
Co061		GAAAACTGAAAGGGAGAAGTGAAAGTGGGAAATTCC	Homo sapiens
Co062	GR/GR	GGGACAAACAGTATGATATGAACACTCAGCTCCATA	Homo sapiens
Co063	PU1/IRF	GGAAATAGAAACCATAGGAAATAGAAACCATAGGAA	Homo sapiens
Co064	CEBP/Ets	TTGTGAAATATATACTTCTGCTTTATATTGTGAAAT	Homo sapiens
Co065	AP1/NFAT	AGGAAAAAATATGTTTCAATAAGGAAAAAATATGTT	Homo sapiens
Co066	AP1/NFkappaB	AAAGAAATTCCAATAAGAGTCATATAAAAGAAATTC	Homo sapiens
Co067		GGAGGAAAAACTGTTTCATACAGAAGGATAGGAGGA	Homo sapiens
Co068	AP1/NFkappaB	AAAGAAATTCCATAAGAGTCAATAAAAGAAATTCCA	Homo sapiens
Co069	NFkappaB/HMGIY	AAAGAAATTCCAATAAAAGAAATTCCAATAAAAGAA	Homo sapiens
Co070	Egr/NFAT	CCCCACCCCATAGGAAAAATACCCCACCCCATAGGA	Homo sapiens
Co071	Ets/HMGIY	AAAAAAAAAAAAAAAAAAAAATAACTTCCTATATTAT	Homo sapiens
Co072	Stat/Stat	TTCTAGGAAATATTCTGATAAATATTCTAGGAAATA	Homo sapiens
Co073		TTCTAGGAAATATTCTGATAAATATTCCATATTCTA	Homo sapiens
Co074		TTCTCAGAAATATTCTGAGAGATATTCTAGGAAATA	Homo sapiens
Co075	AP1/Ets	TGAGTCAATACCCTTCCTGCCATATGAGTCAATACC	Homo sapiens
Co076	AP1/NFAT	TGAGCTAATAGTTTTCCAATATGAGCTAATAGTTTT	Homo sapiens
Co077	NFAT/Oct	TGGAAAATGCAAATATATGGAAAATGCAAATATATG	Homo sapiens
Co078	AP1/GATA	GGCATTCTCTATCTGATTGTTATAATTCATTCCTCA	Homo sapiens

Co079	AP1/Oct	AATTATTCAATAAATTATTCAATAAATTATTCAATA	Homo sapiens
Co080	CEBP/NFkappaB	ACATTGCACAATCTATAGGGATTTTCCATAACATTG	Homo sapiens
Co081		ACATTGCACAATCTATATATCAAATGATAGGGATTT	Homo sapiens
Co082		ACATTGCACAATCTATAGGGATTTTCCCATGATAAC	Homo sapiens
Co083	CEBP/NFkappaB	CAGTTGCAAATCGTGGAATTTCCTATACAGTTGCAA	Homo sapiens
Co084	AP1/NFkappaB	TGACTCAATAGGAATTTCCTATATGACTCAATAGGA	Homo sapiens
Co085	NFkappaB/Stat	TTTCCCCGAAAATAGGGGAATCCCATATTTCCCCGA	Homo sapiens
Co086		TTTCCCCGAAAATAGGGGAATCCCATATTTCCCCGA	Homo sapiens
Co087	PU1/IRF	GGGAAACCGAAAATAGGGAAACCGAAAATAGGGAAA	Homo sapiens
Co088	Ets/Myb	CAGGAAGTATACGGTTTATACAGGAAGTATACGGTT	Homo sapiens
Co089	IRF/NFkappaB	TGGGGATTCCCCACTCCCCTGAGTTTCACTTCTATA	Homo sapiens
Co090		ACCAGCGACTGATATGATGCTAATACTGATTCGTTA	Homo sapiens
Co091	RFX/NFY	CCTAGCAACAGATGATACTGATTGGCCAAAGATACC	Homo sapiens
Co092	AP1/NFY	ATGCGTCAATACTGATTGGCCAAAGATAATGCGTCA	Homo sapiens
Co093	AP1/RFX	CCTAGCAACAGATGCGTCAATACCTAGCAACAGATG	Homo sapiens
Co094		CCTAGCAACAGATGCGTCAATACTGATTGGCCAAAG	Homo sapiens
Co095	Sp1/NFY	TGGGCGGAGTATAGACGAATCAGATATATGGGCGGA	Homo sapiens
Co096	RFX/NFY	CCCAGAAACAAGTGATGAATACAGCCAATGGGATAC	Homo sapiens
Co097	Ebox/Smad	CCTAGACAATACACGTGGATACCTAGACAATACACG	Homo sapiens
Co098	Smad/Smad	GTCTGGACATAGGAGTCAATAGTCTGGACATAGGAG	Homo sapiens
Co099	AP1/ER	AGGTCACGGTGGCCAATATGAATCAATAAGGTCACG	Homo sapiens
Co100	CEBP/NFkappaB	ACACAACTGGGAATAGGGACTTTCCATAACACAACT	Homo sapiens
Co101	AP1/Ets	TGTGTCATTTCCTTATATGTGTCATTTCCTTATATG	Homo sapiens
Co102	AP1/Ets	AGGAAATTAGTCAATAAGGAAATTAGTCAATAAGGA	Homo sapiens
Co103	Ets/AML	GAAGCCACATCCTCTATAGAAGCCACATCCTCTATA	Homo sapiens
Co104	Ets/AML	CAGGATGTGGTTTATACAGGATGTGGTTTATACAGG	Homo sapiens
Co105	Myb/AML	TGTGGTTTATACCGTTAATATGTGGTTTATACCGTT	Homo sapiens
Co106	AP1/NFAT	TGAGCTCAATAGGGTTTCTCCATATGAGCTCAATAG	Homo sapiens
Co107	AP1/CEBP	ATGAGCTCAATAGGGTTTCTCCACCAAGGAAGTTTT	Homo sapiens
Co108	AP1/Ets	TTCCTATAATGAGCTCATATATTCCTATAATGAGCT	Homo sapiens
Co109	AP1/NFkappaB	TGAGCTCAATAGGGTTTCTCCATATGAGCTCAATAG	Homo sapiens
Co110	AP1/CEBPCREB/CREB	TGAGTCAGATAATTTGCTTCAGAATATGAGTCAGAT	Homo sapiens
Co111		TGAGTCAGATATGAGTCAGATATGAGTCAGATATGA	Homo sapiens
Co112	AP1/Pit	AATTCAGTATGAATTTTCAATAGATGCTATATGGGT	Homo sapiens
Co113	AP1/Ets	AGGAAATATGAGTCAATAAGGAAATATGAGTCAATA	Homo sapiens
Co114	Sp1/Oct	GGGCGGGGATAATGCAAATATAGGGCGGGGATAATG	Homo sapiens
Co115	AP1/Ets	AGGAAATGAAGTCAATAAGGAAATGAAGTCAATAAG	Homo sapiens
Co116	AP1/Ets	AGGAAAATATGAATCATATAAGGAAAATATGAATCA	Homo sapiens
Co117	AP1/AP1	TGAAGTCAATATGAATCATATATGAAGTCAATATGA	Homo sapiens
Co118	AP1/Ets	TGACTCAATATCTTCCTTATATGACTCAATATCTTC	Homo sapiens
Co119	E2F/E2F	TTTTCGCGCATATTTGGCGCATATTTTCGCGCATAT	Homo sapiens
Co120		GGGTTTCCCCATAGGGATTTCCCATAACTTTCTATT	Homo sapiens
Co121	NFkappaB/NFkappaB	GGGTTTCCCCATAGGGATTTCCCATAGGGTTTCCCC	Homo sapiens
Co122	CEBP/NFkappaB	ATTGCATAGGAAATTCCGATAATTGCATAGGAAATT	Homo sapiens
Co123	Sp1/Stat	TTTCCGGGAAAATACCGCCCATATTTCCGGGAAAAT	Homo sapiens
Co124	AP1/NFkappaB	TGACTCTATAGGGTTTTCCATATGACTCTATAGGGT	Homo sapiens
Co125	NF1/Sp1	GTCATGGCGACTGTCCATAGTCATGGCGACTGTCCA	Homo sapiens
Co126	CEBP/AML	ATTTCCAAAATATGTGGTATAATTTCCAAAATATGT	Homo sapiens
Co127	Ets/AML	TGTGGTATAGGGGAAATATGTGGTATAGGGGAAATA	Homo sapiens
Co128	Sp1/Smad	GGGCGGATAATGCAGACAATAGGGCGGATAATGCAG	Homo sapiens
Co129	Sp1/Smad	GGGCGGATAATGCAGACAATAGGGCGGATAATGCAG	Homo sapiens
Co130	AhR/Sp1	CCCCGCCCATACACGCCGGCCGGATACCCCGCCCAT	Homo sapiens
Co131	ER/Sp1	GGGCAATAGGCGGGATAGGGCAATAGGCGGGATAGG	Homo sapiens
Co132	AP1/Ets	AGCGGATGTGATATGAGTCAATAAGCGGATGTGATA	Homo sapiens
Co133	COUP/HNF4	ACGTGACCTTGGGGGACGTCATTATACTGTTGGCCA	Homo sapiens
Co134	Oct/Stat	ATTTGCATTTCTATGAAATAATTTGCATTTCTATGA	Homo sapiens
Co135	IRF/STAT	GGTTTCAGTTTTCCATATTCCTGTAAATAGGTTTCA	Homo sapiens
Co136	AP1/NFAT	AGGAAAATAAACTACAATAAGGAAAATAAACTACAA	Homo sapiens
Co137	SREBP/Sp1	ATCACCCCACATACCTCCCCCTGCATAATCACCCCA	Homo sapiens
Co138		GAAAATTTCCATAGAAAATTTCCATAGAAAATTTCC	Homo sapiens
Co139	IRF/NFAT	GAAAATTTCCATAGAAAATTTCCATAGAAAATTTCC	Homo sapiens
Co140	CEBP/NFkappaB	GTGATGTAAATAGGGACACTCCATAGTGATGTAAAT	Homo sapiens
Co141		GGGGGTGACCCCATAGGGGGTGACCCCATAGGGGGT	Homo sapiens
Co142	IRF/NFkappaB	GTTTTCTTTTCCATAGGGGATGCCCCATAGTTTTCT	Homo sapiens
Co143	RFLAT/NFkappaB	TTTTGGAAACTCCCCTTAGGGGATGCCCCATATTTT	Homo sapiens
Co144	Egr/NFAT	GTGGGCGGAAACTTATAGTGGGCGGAAACTTATAGT	Homo sapiens
Co145		AGAAACGGAGGATAGGGGCGGGGCGCGATATGCGTC	Homo sapiens
Co146		TTCCATATTTCAAAGATATGAGTCAATATTCCATAT	Homo sapiens
Co147	Ets/HMG	TTCCATATTTCAAAGATATTCCATATTTCAAAGATA	Homo sapiens
Co148		ATTATGGGAAACCATAATTATGGGAAACCATAATTA	Homo sapiens
Co149	Ets/Myb	CCGTTATCATAGGATATACCGTTATCATAGGATATA	Homo sapiens
Co150	SF1/Sp1	CCGCCCCATATATCCTTGACATACCGCCCCATATAT	Homo sapiens
Co151	PDX/HNF3	CTTTAATTGGTTATACAGCCTTTTTTGTTTATTTAT	Homo sapiens
Co152	Sp1/NFY	CCCGCCCCATAATTGGATACCCGCCCCATAATTGGA	Homo sapiens
Co153	Sp1/Ets	ACAGGAATATACTCGCCCATAACAGGAATATACTCG	Homo sapiens
Co154	Ebox/Stat	TTCTGATAAAATACACGTGATATTCTGATAAAATAC	Homo sapiens
Co155		TTCTGATAAAATACACGTGATAGAAAGTGAAAGGAT	Homo sapiens
Co156	CEBP/Ets	CCAATATAGAGGAAATACCAATATAGAGGAAATACC	Homo sapiens
Co157	CEBP/NFkappaB	GAAATTCCCCATAATGTTGCAAATAGAAATTCCCCA	Homo sapiens
Co158	Sp1/Stat	ATCACCCCACATAATTGGCATAATCACCCCACATAA	Homo sapiens
Co159	SF1/Sox	TCTTTGAGAATACCAAGGTCGCATATCTTTGAGAAT	Homo sapiens
Co160	Sp1/Smad	GGGGGCGGATAGCCTATAGGGGGCGGATAGCCTATA	Homo sapiens

7. Bibliography

Books and Papers:

Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P: Molecular biology of the cell. 4. Edition; Garland Science, New York 2002

Allison LA: Fundamental molecular biology. 2. Edition; John Wiley & Sons, Malden 2011

Biedler JL, Roffler-Tarlov S, Schachner M, Freedman LS (1978): Multiple neurotransmitter synthesis by human neuroblastoma cell lines and clones. Cancer Res 38(11 Pt 1), [3751-7]

Birnboim HC, Doly J (1979): A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res 7(6), 1513-23

Botvinnik A, Wichert SP, Fischer TM, Rossner MJ (2010): Integrated analysis of receptor activation and downstream signaling with EXTassays. Nat Methods 7(1), 74-80

Brivanlou AH, Darnell JE Jr. (2002): Signal transduction and the control of gene expression. Science 295(5556), 813-8

Bronstein I, Fortin J, Stanley PE, Stewart GS, Kricka LJ (1994): Chemiluminescent and bioluminescent reporter gene assays. Anal Biochem 219(2), 169-81

Chanda SK, Caldwell JS (2003): Fulfilling the promise: drug discovery in the post-genomic era. Drug Discov Today 8(4), 168-74

Chytil M, Verdine GL (1996). The Rel family of eukaryotic transcription factors. Curr Opin Struct Biol. 6(1):91-100

Darnell JE Jr (2002): Transcription factors as targets for cancer therapy. Nat Rev Cancer. 2(10):740-9

Deplancke B, Dupuy D, Vidal M, Walhout AJ (2004): A gateway-compatible yeast one-hybrid system. Genome Res 14(10B), 2093-101

Fogh J, Wright WC, Loveless JD (1977): Absence of HeLa cell contamination in 169 cell lines derived from human tumors. J Natl Cancer Inst 58(2), 209-14

Giard DJ, Aaronson SA, Todaro GJ, Arnstein P, Kersey JH, Dosik H, Parks WP (1973): In vitro cultivation of human tumors: establishment of cell lines derived from a series of solid tumors. J Natl Cancer Inst 51(5), 1417-23

Graham FL, Smiley J, Russell WC, Nairn R (1977): Characteristics of a human cell line transformed by DNA from human adenovirus type 5. J Gen Virol 36(1), 59-74

Greene LA, Tischler AS (1976): Establishment of a noradrenergic clonal line of rat adrenal pheochromocytoma cells which respond to nerve growth factor. Proc Natl Acad Sci 73(7), 2424-8

Hanahan D, Weinberg RA (2000): Hallmarks of cancer. Cell 100(1), 57-70

Hanahan D, Weinberg RA (2011): Hallmarks of cancer: The Next Generation. Cell 144(5), 646-74

Hartley JL, Temple GF, Brasch MA (2000): DNA cloning using in vitro site-specific recombination. Genome Res 10(11), 1788-95

Heinemeyer T, Wingender E, Reuter I, Hermjakob H, Kel AE, Kel OV, Ignatieva EV, Ananko EA, Podkolodnaya OA, Kolpakov FA (1998): Databases on transcriptional regulation: TRANSFAC, TRRD and COMPEL. Nucleic Acids Res 26(1), 362-7

John LB, Ward AC (2011): The Ikaros gene family: transcriptional regulators of hematopoiesis and immunity. Mol Immunol 48(9-10), 1272-8

Karin M, Liu Zg, Zandi E (1997): AP-1 function and regulation. Curr Opin Cell Biol 9(2), 240-6

Kibbe WA (2007): OligoCalc: an online oligonucleotide properties calculator. Nucleic Acids Res 35(Web Server issue), W43-6

Ko CH, Takahashi JS (2006). Molecular components of the mammalian circadian clock. Hum Mol Genet. 15 Spec No 2:R271-7

Lee KA, Masson N (1993). Transcriptional regulation by CREB and its relatives. Biochim Biophys Acta. 1174(3):221-33

Leibovitz A, Stinson JC, McCombs WB 3rd, McCoy CE, Mazur KC, Mabry ND (1976): Classification of human colorectal adenocarcinoma cell lines. Cancer Res 36(12), 4562-9

Levine M, Tjian R (2003): Transcription regulation and animal diversity. Nature 424(6945), 147-51

Li X, Jiang X, Yaoi T (2006): High throughput assays for analyzing transcription factors. Assay Drug Dev Technol 4(3), 333-41

Luo Y, Batalao A, Zhou H, Zhu L (1997): Mammalian two-hybrid system: a complementary approach to the yeast two-hybrid system. Biotechniques 22(2), 350-2

Maniatis T, Goodbourn S, Fischer JA (1987): Regulation of inducible and tissue-specific gene expression. Science 236(4806), 1237-45

McAllister RM, Isaacs H, Rongey R, Peer M, Au W, Soukup SW, Gardner MB (1977): Establishment of a human medulloblastoma cell line. Int J Cancer 20(2), 206-12

Mullis KB (1990): Target amplification for DNA analysis by the polymerase chain reaction. Ann Biol Clin 48(8), 579-82

Mullis KB, Faloona FA (1987): Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. Methods Enzymol. 155, 335-50

Noguchi P, Wallace R, Johnson J, Earley EM, O'Brien S, Ferrone S, Pellegrino MA, Milstien J, Needy C, Browne W (1979): Characterization of the WIDR: a human colon carcinoma cell line. In Vitro 15(6), 401-8

Oeckinghaus A, Ghosh S (2009). The NF-kappaB family of transcription factors and its regulation. Cold Spring Harb Perspect Biol. 1(4):a000034

Pontén J, Saksela E (1967): Two established in vitro cell lines from human mesenchymal tumours. Int J Cancer 2(5), 434-47

Rao A, Luo C, Hogan PG (1997): Transcription factors of the NFAT family: regulation and function. Annu Rev Immunol 15, 707-47

Rasheed S, Nelson-Rees WA, Toth EM, Arnstein P, Gardner MB (1974): Characterization of a newly derived human sarcoma cell line (HT-1080). Cancer 33(4), 1027-33

Rebollo A, Schmitt C (2003): Ikaros, Aiolos and Helios: transcription regulators and lymphoid malignancies. Immunol Cell Biol 81(3), 171-5

Romanov S, Medvedev A, Gambarian M, Poltoratskaya N, Moeser M, Medvedeva L, Gambarian M, Diatchenko L, Makarov S (2008): Homogeneous reporter system enables quantitative functional assessment of multiple transcription factors. Nat Methods 5(3), 253-60

Rossner MJ, Oster H, Wichert SP, Reinecke L, Wehr MC, Reinecke J, Eichele G, Taneja R, Nave KA (2008): Disturbed clockwork resetting in Sharp-1 and Sharp-2 single and double mutant mice. PLoS One 3(7), e2762

Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M, Leamon JH, Johnson K, Milgrew MJ, Edwards M (2011): An integrated semiconductor device enabling non-optical genome sequencing. Nature 475(7356), 348-52

Rydholm H, Boström S, Eriksson E, Risberg B (1995): Complex intracellular signal transduction regulates tissue plasminogen activator (t-PA) and plasminogen activator inhibitor type-1 (PAI-1) synthesis in cultured human umbilical vein endothelium. Scand J Clin Lab Invest 55(4), 323-30

Sanger F, Nicklen S, Coulson AR (1977): DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci 74(12), 5463-7

Scherer WF (1954): Current applications of tissue cultures in medicine. J Lancet 74(2), 52-8

Shaulian E, Karin M (2002): AP-1 as a regulator of cell life and death. Nat Cell Biol 4(5), E131-6

Semenza GL, Agani F, Booth G, Forsythe J, Iyer N, Jiang BH, Leung S, Roe R, Wiener C, Yu A (1997). Structural and functional analysis of hypoxia-inducible factor 1. Kidney Int. 51(2):553-5

Siletz A, Schnabel M, Kniazeva E, Schumacher AJ, Shin S, Jeruss JS, Shea LD (2013): Dynamic transcription factor networks in epithelial-mesenchymal transition in breast cancer models. PLoS One 8(4):e57180

Smale ST, Kadonaga JT (2003): The RNA polymerase II core promoter. Annu Rev Biochem 72, 449-79

Soule HD, Vazguez J, Long A, Albert S, Brennan M (1973): A human cell line from a pleural effusion derived from a breast carcinoma. J Natl Cancer Inst 51(5), 1409-16

Stagljar I, Korostensky C, Johnsson N, te Heesen S (1998): A genetic system based on split-ubiquitin for the analysis of interactions between membrane proteins in vivo. Proc Natl Acad Sci 95(9), 5187-92

Suardet L, Gaide AC, Calmès JM, Sordat B, Givel JC, Eliason JF, Odartchenko N (1992): Responsiveness of three newly established human colorectal cancer cell lines to transforming growth factors beta 1 and beta 2. Cancer Res 52(13), 3705-12

Tantin D, Schild-Poulter C, Wang V, Haché RJ, Sharp PA (2005): The octamer binding transcription factor Oct-1 is a stress sensor. Cancer Res 65(23), 10750-8

Towbin H, Staehelin T, Gordon J (1992): Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. Biotechnology 24, 145-9

Weis J, Fine SM, David C, Savarirayan S, Sanes JR (1991): Integration site-dependent expression of a transgene reveals specialized features of cells associated with neuromuscular junctions. J Cell Biol 113(6), 1385-97

Wingender E, Dietze P, Karas H, Knüppel R (1997): TRANSFAC: a database on transcription factors and their DNA binding sites. Nucleic Acids Res 24(1), 238-41

Wingender E, Schoeps T, Dönitz J (2013): TFClass: an expandable hierarchical classification of human transcription factors. Nucleic Acids Res 41(Database issue), D165-70

Xie X, Lu J, Kulbokas EJ, Golub TR, Mootha V, Lindblad-Toh K, Lander ES, Kellis M (2005): Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals. Nature 434(7031), 338-45

Yamazaki S, Numano R, Abe M, Hida A, Takahashi R, Ueda M, Block GD, Sakaki Y, Menaker M, Tei H (2000): Resetting central and peripheral circadian oscillators in transgenic rats. Science 288(5466), 682-5

Yaniv M (1984): Regulation of eukaryotic gene expression by transactivating proteins and cis acting DNA elements. Biol Cell 50(3), 203-16

Yeh JE, Toniolo PA, Frank DA (2013): Targeting transcription factors: promising new strategies for cancer therapy. Curr Opin Oncol. 25(6):652-8

Yoo SH, Yamazaki S, Lowrey PL, Shimomura K, Ko CH, Buhr ED, Siepka SM, Hong HK, Oh WJ, Yoo OJ (2004): PERIOD2: LUCIFERASE real-time reporting of circadian dynamics reveals persistent circadian oscillations in mouse peripheral tissues. Proc Natl Acad Sci 101(15), 5339-46

Zheng R, Blobel GA (2010): GATA Transcription Factors and Cancer. Genes Cancer 1(12), 1178-88

Webpages:

http://www.who.int/topics/cancer/en/ ; release on 10.06.2015

Acknowledgements

This work was done in the Research Group "Gene Expression" at the Max-Planck-Institute for Experimental Medicine under the supervision of Prof. Dr. Moritz Rossner.
I'm grateful to my supervisor for giving me the opportunity to work on this project. I deeply appreciate his invaluable input to this project, his guidance, patience and support.
I thank him for being always available and open for a discussion and also for the scientific knowledge and experience.

I furthermore thank Prof. Dr. Klaus-Armin Nave, for giving me the great opportunity to work in his department.

I thank Dr. Gabriela Salinas-Riester and Lennart Opitz for the help with microarray experiments.
A lot of thanks to Dr. Sven Wichert not only for the Next Generation Sequencing analysis work that he has done for this project but also for his jokes and the fun working together.

I'm grateful to my colleagues Alexandra, Alexander, Jessica, Magda, Lisa, Sven, Wilko, Ali, Dorota, Sabrina and Carolin for creating a nice working atmosphere, for their technical input and great team work.
Many thanks go to all the other members of the Neurogenetics department in particular to Ulli, Martin, Gabriele and Michaela for being so nice and always helpful.

I'm very grateful to Moritz for proofreading the thesis and for the help at the last stages of writing and data analysis.