

“Combining induced protease fragment assembly and
microarray analysis to monitor signaling in living cells”

PhD Thesis

in partial fulfillment of the requirements
for the degree “Doctor of Philosophy (PhD)”
in the Molecular Biology Program
at the Georg August University Göttingen,
Faculty of Biology

Submitted by
Anna Botvinnik

Born in
Vinnytsia (Ukraine)

2009

Declaration

I hereby declare that the PhD thesis entitled, “**Combining induced protease fragment assembly and microarray analysis to monitor signaling in living cells**”, has been prepared independently and with no other sources and aids than quoted. I would like to gratefully acknowledge Sven Wichert for the microarray data analysis, Dr. Reiner Hitt for generation of self-spotted microarrays and Dr. Fritz Benseler for the synthesis of the complex oligonucleotide library.

Anna Botvinnik

Göttingen, 30.04.2009

Acknowledgements

The following work was done in the Research Group “Gene Expression” at the Max-Planck-Institute for Experimental Medicine under the supervision of 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 knowledge and experience both, scientific and non scientific that I gained from him.

I would like to express my gratitude to the members of my doctoral committee Prof. Klaus-Armin Nave, Prof. Frauke Melchior and Prof. Ernst Wimmer for their support and advice at the different stages of the project and their valuable time and attention.

I thank Dr. Reiner Hitt, Dr. Gabriela Salinas-Reister and Lennart Opitz for the help with microarray experiments.

A lot of thanks to Sven Wichert not only for the microarray 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 Magda, Minou, Lisa, Sven, Wilko, Toby and Carolin for creating a nice working atmosphere, for their technical input and great team work. Also thanks to Eva and Katrin who have complemented our group recently. I would like particularly to thank Toby for creative discussions and Carolin for her expertise in microarray protocols.

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.

It was great to have my fellow friends Foteini, Viki, Olga, Schanila, Amit, Patricia, Bhanva, Andrea and Alena that made the time in Göttingen so much nicer. I appreciate and value this friendship and hope it will survive much longer than our PhD time.

I owe many thanks to Harry Scherer for being a good friend and Hajo Horn for helping me out with notebook explosions and many other computer related problems.

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

I also thank Schanila for correcting the materials and methods section.

I would like to thank the organization committee of the IMPRS, especially Dr. Steffen Burkhardt and Ivana Bacakova for help in every possible situation.

I am grateful to my family and my husband Sergey for the sacrifices he had to make in the name of science, for his support and constant attention throughout this time.

Table of Contents

<i>Acknowledgements</i>	i
List of figures	v
Abbreviations	vi
1. ABSTRACT.....	1
2. INTRODUCTION.....	2
<i>Objective</i>	2
<i>High throughput technologie</i>	3
<i>DNA microarrays</i>	4
<i>Barcode arrays: novel applications of the established technology</i>	5
<i>Reporter gene assay</i>	6
<i>Multiplexed reporter gene assays</i>	7
<i>Split-TEV assay</i>	8
<i>Other protein complementation assays</i>	9
<i>ErbB signaling pathway</i>	10
<i>Aim of the study</i>	14
3. MATERIALS.....	15
Laboratory supplies and equipment	15
Reagents	16
Molecular biology solutions	17
Microarray related buffers and solutions	19
Luciferase assay buffers	20
Solutions for cell culture	20
Mammalian cell lines	21
Bacterial Strains	21
List of Vectors and Plasmids	22
List of Primers	23
4. METHODS.....	25
<i>Molecular biology methods</i>	25
Molecular cloning	25
Gateway Recombination-mediated cloning	25
Multisite Gateway Recombination mediated cloning	26
PEG-mediated DNA precipitation	28
Determination of nucleic acids concentration	28
DNA sequencing and Oligonucleotide synthesis	29
Transformation of chemically competent bacteria	29
Electroporation of bacteria	29
Isolation of plasmid DNA	30
Generating frozen stocks of bacterial cultures	30

Polymerase Chain Reaction	30
Designing primer sequences	31
Colony PCR	31
Agarose gel electrophoresis	32
Phenol Chloroform extraction and DNA precipitation	32
<i>Cell biology Methods</i>	33
Culturing of eukaryotic cells	33
Passaging of eukaryotic cell lines	33
Thawing of eukaryotic cell lines	33
Generating frozen stocks of eukaryotic cells	33
Transfection of mammalian cells	34
Transfection in solution	34
Luciferase reporter gene assay	35
<i>Preparation of the microarray target</i>	36
RNA isolation	36
RNA precipitation	36
First Strand cDNA synthesis	37
Quantitative real-time PCR	37
In Vitro Transcription	37
Purification of short RNA	38
Labeling of the microarray target	38
<i>Microarray hybridization and analysis</i>	39
Hybridization of Agilent microarrays	
Hybridization of Agilent microarrays	39
Hybridization of NimbleGen microarrays	39
Hybridization of the self spotted microarrays	39
Microarray Scanning and Data Extraction	40
Microarray Data Analysis	40
<i>Synthesis and cloning of the EXT reporter library</i>	40
EXT library synthesis	40
Cloning of the pENTR EXT library	41
Cloning of the G5 EXT library	41
Cloning of the <i>Cis</i> -element EXT reporter constructs	42
Transfections for Microarray Experiments	42
List of <i>Cis</i> -elements and associated EXT reporters	43
5. RESULTS	45
<i>The principle of EXT assays</i>	45
<i>Structure and properties of the EXT reporter library</i>	47
<i>Optimization of the hybridization conditions</i>	49
<i>EXT performance in microarray hybridization</i>	50
<i>Dynamic range of the microarray based EXT measurement</i>	58
<i>Robustness against amplification biases</i>	60
<i>Design and analysis of EXT reporter constructs</i>	62
<i>Comparison of the "RNA" and luciferase reporter</i>	66
<i>Multiplexed EXT analysis with microarrays</i>	68
<i>Monitoring of ErbB receptor activation and downstream signaling</i>	71
<i>Improved design of the EXT reporter constructs</i>	79

6. DISCUSSION.....	81
<i>EXTassay readout characteristics</i>	81
<i>EXTassay and standard reporter gene assays</i>	82
<i>The relevance and scalability of the EXTassay</i>	83
<i>Flexibility of the EXTassay</i>	84
<i>ErbB signaling measurement with EXTassays</i>	87
7. CONCLUSIONS.....	89
8. Curriculum Vitae.....	90
9. BIBLIOGRAPHY.....	92

List of figures

Figure 1. The principle of the split-TEV assay	9
Figure 2. Domain structure and dimerization mechanism of the ErbB receptors	11
Figure 3. An overview of the Neuregulin/ErbB signaling	12
Figure 4. Schematic representation of the three-fragment recombination cloning	27
Figure 5. The principle of the EXT assay	46
Figure 6. Structure and synthesis strategy of the EXT oligonucleotide library	48
Figure 7. Melting temperature and similarity profiles of EXT oligonucleotide libraries	48
Figure 8. Control EXTs c1-5	50
Figure 9. Cross hybridization profile for single base pair mismatch controls	53
Figure 10. Cross hybridization profile for two base pair mismatch controls	54
Figure 11. Comparison of the cross hybridization profiles for 4 and 8 base versus 1 or 2 words mismatches at corresponding positions	55
Figure 12. Combining single central mismatches with single word mismatches	56
Figure 13. Cross-hybridization profiles with multiple word mismatches	57
Figure 14. Assessment of the dynamic range of microarray based EXT analysis	59
Figure 15. Controlling a potential PCR amplification bias	61
Figure 16. EXT amplification and subcloning strategy	63
Figure 17. EXT reporter constructs	65
Figure 18. Stability of the Protein-based versus mRNA-based reporter	67
Figure 19. Testing EXT assay performance under control conditions	69
Figure 20. Analysis of the “GV” control EXT assay	71
Figure 21. Experimental steps of the EXT analysis from transfections to microarray hybridization	72
Figure 22. Split-TEV assays integrated into the analysis of the ErbB signaling pathway with EXT assays	73
Figure 23. Luciferase measurements of the Split-TEV assays to monitor ErbB receptor interactions	75
Figure 24. Kinetic measurement of the Nrg1/ErbB receptor signaling with EXT assays	77
Figure 25. Improved design for EXT reporter constructs	80

Abbreviations

ATCC	American Tissue Culture Collection
aaRNA	aminoallyl-modified RNA
aaUTP	aminoallyl-UTP
AmpR	Ampicillin resistance
ATP	Adenosine triphosphate
cAMP	Cyclic 3'5'-adenosine monophosphate
cDNA	complementary DNA
<i>Cis-</i>	a prefix meaning "on the same side"
CMV	Cytomegalovirus
Dec	"Decoder" primer for EXT amplification
DMEM	Dulbecco's modified Eagle Medium
DMSO	Dimethyl sulfoxide
Dox	Doxycycline
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
dNTP	Deoxyribonucleotide triphosphate
DTT	1,4-Dithiothreitol
<i>E.coli</i>	<i>Escherichia coli</i>
EDTA	Ethylendiaminetetraacetate
EGF	Epidermal Growth Factor
et al.	and others
EtOH	Ethanol
EYFP	Enhanced Yellow Fluorescent Protein
EXT	<i>Expressed unique sequence tag</i>
FBS	Fetal Bovine Serum
f.c.	final concentration
G5	five times Gal4 binding site (or 5x upstream activation sequence)
Grb2	Growth factor Receptor-Bound protein 2
GV	Gal4-VP16 synthetic transcriptional activator
HS	Horse Serum
IVT	<i>in vitro</i> transcription
KanR	Kanamycin resistance
LB	Luria and Bertani medium (or Luria's broth)
MAPK	Mitogen-activated Protein Kinase
MES	2-(N-morpholino)ethanesulfonic acid
mid-prep	middle scale plasmid DNA preparation
mini-prep	small scale plasmid DNA preparation
mRNA	messenger-RNA
NCBI	<i>National Center for Biotechnology Information</i>
NGF	Nerve Growth Factor
NLS	Nuclear Localization Sequence
Nrg1	neuregulin 1
Oligo	Oligonucleotide
ON	Overnight
pA	polyadenylation signal
PBS	Phosphate-buffered Saline
PCR	Polymerase Chain Reaction
PEG	Polyethylene glycole

pH	Negative decimal logarithm of the hydrogen ions (H ⁺)
PI3K	Phosphoinositide 3-kinase
PLL	Poly-L-Lysin
RLUs	Relative Luciferase Units
RNA	ribonucleic acid
RNase	Ribonuclease
RT	Room Temperature
SDS	Sodium Dodecyl Sulfate
SHC	Src Homology 2 Domain Containing Transforming Protein
SOC	Super Optimal Broth medium with Catabolite repression
SV40	Simian Virus 40
TAE	Tris-Acetate-EDTA
Taq	<i>Thermus aquaticus</i>
TATA	TATA-box
TB	Terrific Broth
TEV	Tobacco Etch Virus NIa protease
TK	Thymidinkinase promoter (from herpes simplex virus)
T _m	Melting temperature
Tris	Tris(hydroxymethyl)aminomethane
UTP	Uridine triphosphate
UV	Ultraviolet

Units of measurement:

A	absorbance
g	gram
b	bases
bp	base pairs
°C	degrees Celsius
h	hour
kb	kilobase
kD	kilo dalton
l	liter
M	molar
min	minute
n	number
OD	optical density
rpm	rotation per minute
sec	seconds
U	units of enzymatic activity
V	volume
w/v	weight / volume

DNA and RNA bases:

A	adenine
G	guanine
T	thymine
C	cytosine
U	uracil
W	A or T
S	G or C
N	any base

Power prefixes:

m	milli (10 ⁻³)
μ	micro (10 ⁻⁶)
n	nano (10 ⁻⁹)
p	pico (10 ⁻¹²)
f	femto (10 ⁻¹⁵)

ABSTRACT

The ability to monitor multiple signaling events simultaneously in living cells is essential to better understand complex biological processes. So far, DNA-microarray technologies provide global scale data mainly restricted to the level of gene expression. This information is not sufficient to understand the upstream regulatory mechanisms that lead to gene expression changes. Most proteomic technologies also provide large scale measurement but usually depend on in vitro synthesized peptides or require biochemical manipulations. High throughput technologies are required for functional characterization and monitoring of signaling components in living cells.

Here, an experimental approach is presented termed EXTassay that enables quantitative and parallel measurements of various signaling events upstream of mRNA expression. EXTassay incorporates various cellular assays that are coupled to reporter gene expression. To achieve multiplexing, we have generated a complex and optimized library of short expressed oligonucleotide tags (EXTs). Each unique EXT can replace a classical reporter gene and serves as a unique identifier for tracking and quantification of a defined cellular assay. Multiple EXT-reporters expressed in the same cell or cell population can be isolated and analyzed by custom microarray hybridization. We have established protocols and optimized the microarray readout for reliable EXT quantification. We applied the EXTassay to analyze the neuregulin 1 induced ErbB receptor tyrosine kinase signaling in PC-12 cells. We used transcriptionally coupled split-TEV protein complementation assays to monitor ErbB receptor dimerization and phosphorylation dependent interaction with downstream signaling proteins. In addition, we employed 30 different *cis*-regulatory elements to assess the downstream signaling. All assays were coupled to unique EXTs and analyzed by microarrays. By analyzing three different receptor complexes (ErbB2/2, 2/3 and 2/4), we were able to measure receptor specific differential signaling effects and demonstrate that EXTassays can be applied for the quantitative profiling of activated signaling pathways.

INTRODUCTION

Objective

To maintain homeostasis cells adapt continuously to changing environmental conditions. Genetic and environmental factors can lead to a progressive aberration of this adaptive behavior ([Fishman and Porter 2005](#)). One example for such a process is the development of cancer. According to the current state of research cancer cells lose progressively diverse mechanisms that inhibit uncontrolled proliferation. The interaction of these processes, however, has not been properly understood ([Araujo, Liotta et al. 2007](#)). Other examples comprise genetic complex and often chronic psychiatric disorders ([Agid, Buzsaki et al. 2007](#)). The complex genetic disorders as well as the simple monogenic (mendelian) disorders have in common that the dynamic equilibrium of physiologic processes in the cell changes at a certain time. Only an improved understanding of all cellular processes can lead to the development of new therapies and new pharmaceutical agents ([Fishman and Porter 2005](#)). The advent of modern genomic and proteomic technologies has dramatically changed the current concepts of molecular biology as exemplified by the paradigm shift in cellular signaling. Linear cascade models evolved towards highly interconnected signaling networks operating at multiple levels e.g. offering new perspective on drug discovery opportunities ([Papin, Hunter et al. 2005](#); [Kitano 2007](#); [Hopkins 2008](#)). Lessons learned from drug discovery show that focusing on one or several proteins that have been proven to play an important role in particular disease might not be sufficient to predictably modulate the behavior of the system ([Kitano 2007](#)). The findings of systems biology showed that the normal or pathologic behaviour of cells has to be understood in the context of the interaction of all cellular components ([Friedman and Perrimon 2007](#); [Perrimon, Friedman et al. 2007](#)).

Fast development of novel technologies gives rise to highly multiplexed experimental approaches that deal not only with gene expression changes but with a variety of cellular assays to gain an insight into signal transduction mechanisms and other cellular functions. Such assays hold a promise for more successful identification of drug targets and more efficient screening for therapeutic compounds ([Chanda and Caldwell 2003](#)).

High throughput technologies

Collection of comprehensive data sets with standard techniques of biochemistry and cell biology is very tedious and cost intensive. Moreover large scale data sets collected so far are mostly based on *in vitro* analysis ([Natarajan, Lin et al. 2006](#)).

Several high throughput methods can be used to analyze cellular functions ([Chanda and Caldwell 2003](#)). Developments in high throughput microscopy and image analysis have opened a way for large dataset collections using well established cellular assays based on fluorescent reporters. High-throughput microscopy has been successfully used to e.g. monitor the dynamics of the subcellular distribution of GFP-fusion constructs ([Starkuviene, Liebel et al. 2004](#)), to quantify GFP-complementation based signaling sensors in living cells ([MacDonald, Lamerdin et al. 2006](#)) or to analyze transfected cell arrays ([Bailey, Wu et al. 2002](#)). Flow cytometry is another technique often used for high throughput quantification of fluorescent reporters or FRET assays ([Gertz and Cohen 2009](#)). Fixation protocols for quantitative immunocytochemical analyses have been combined with color-barcoding schemes to achieve multiplexing ([Krutzik and Nolan 2006](#)). Multiplexed RNAi screening strategies are being used to determine loss of function phenotypes on the large scale ([Schlabach, Luo et al. 2008](#)). Protein-interaction measurements with two-hybrid assays have been multiplexed using yeast-arrays ([Suter, Fetchko et al. 2007](#); [Suter, Kittanakom et al. 2008](#)).

Protein arrays are commonly used in high throughput proteomics to collect large datasets on protein-protein interaction. This technique is a pure *in vitro* approach that uses antibodies or spotted peptides obtained by *in vitro* translation. Such approach might not be informative for functional assessments of full length transmembrane proteins or transiently modified signaling proteins. Identification of protein composition by mass spectrometry is another powerful proteomics technology that works *in vitro* with extracted proteins and is not capable to provide dynamic information about the functional state of full length proteins in their native cellular environment.

So far, most techniques are restricted to a particular biochemical reaction or class of molecules and thus can only provide an isolated picture of the molecularly very different signaling events that are associated with a particular stimulus.

DNA microarrays

The entirety of transcripts, i.e. those parts of the genome that are used for the synthesis of mRNA, is known as the “transcriptome”. It is far more dynamic and complex than thought a few years ago. The most advanced technology for the high throughput screening of the transcriptome at the moment is the “microarray” technology ([Chanda and Caldwell 2003](#)). Microarrays rely on the specific hybridization of nucleic acids in solution (targets) with complementary nucleic acids bound to a solid surface (probes). A reliable analysis of almost the complete transcriptome is possible due to the sequence-specific recognition between unique probes and corresponding targets. The first microarrays were applied for gene expression analysis in 1995 with only 45 genes ([Schena, Shalon et al. 1995](#)). Since then, the microarray technique has made a tremendous progress and became a widely used and very powerful technology that made invaluable contributions for functional genomics. Three different approaches to microarray manufacturing were developed simultaneously: inkjet printing (by Edwin Southern) and photolithography (by Stephen Fodor) are two different approaches for the de novo synthesis of oligonucleotides on the glass surface ([Fodor, Rava et al. 1993](#)) and an alternative technique developed by Patrick Braun based on spotting readily available PCR products with the help of tiny pins. First whole genome microarrays were performed to analyze gene expression in yeast using the spotting technology ([DeRisi, Iyer et al. 1997](#)). Recently, a modification of photoactivated in situ oligo synthesis technology was introduced using electronically manipulated micromirrors of the digital light processor device instead of photolithographic masks ([Nuwaysir, Huang et al. 2002](#)). All platforms have proven to be reliable and useful for gene expression studies and many other applications. However preparation of self-spotted microarrays is time consuming and does not guarantee reproducible quality from array to array. Microarray based transcriptomics continues to generate new data unraveling novel phenomena in cellular function and disease mechanisms. Microarray technology has enabled not only genome scale transcriptomic research but also other kinds of high throughput assays such as chromatin immunoprecipitation, single nucleotide polymorphism, profiling of small non coding RNA expression and other ([Zhao, Yin et al. 2007](#)). Alternatively, high throughput sequencing techniques may complement

microarray approaches for globally scaled analyses of the genome and transcriptome in the near future ([Shendure 2008](#); [Shendure, Porreca et al. 2008](#)).

Barcode arrays: novel applications of the established technology

Microarray technologies are used for sequence specific detection and quantification of DNA and RNA samples not only from natural biological samples but also from synthetic DNA libraries. Such DNA sequences are used as unique identifiers (barcodes or tags) linked to other molecules such as siRNAs or targeting vectors to generate tagged null alleles via recombination in yeast. Multiple DNA barcodes are quantified by hybridization to a diagnostic microarray or by high throughput sequencing ([Mannocci, Zhang et al. 2008](#)). DNA-based molecular barcodes analyzed by microarray hybridization have proven to be useful in several applications. A library of yeast deletion mutants was constructed covering 96% of all annotated *Saccharomyces cerevisiae* open reading frames where each deleted gene was replaced by a barcode DNA sequences for tracking of individual deletion strains. This enabled quantitative survival assay for all the mutants grown in one culture by measurement of the relative abundance of the unique DNA barcodes ([Giaever, Chu et al. 2002](#)). The technique was also applied to characterize drug sensitivities of tagged heterozygote strains and to study the interaction of various drugs with certain genetic backgrounds ([Giaever, Shoemaker et al. 1999](#)). Such approaches are helpful for the identification of drug targets or mechanisms of their action ([Lum, Armour et al. 2004](#)). In another “barcode” application, DNA sequences were used in combination with specific shRNAs targeting up to 20 000 different genes in human mammary epithelial cell lines MCF-10A and MDF-MB435. The barcode sequences were used in a fashion similar to the yeast approach to quantify the impact of specific shRNAs on cell survival ([Silva, Marran et al. 2008](#)). Unique DNA sequences can also serve for measurement of multiple cellular assays within one cell. The authors of a recent review suggested using unique oligonucleotides as specific reporter genes for the measurement of multiple transcription factor activities and for parallel assessment of transcription factor DNA binding in vitro and for identification of transcription factor complexes with pull down assays ([Li, Jiang et al. 2006](#)).

Reporter gene assays

A reporter gene typically encodes a protein or enzyme which is induced by a cellular process of interest and which can be easily detected and quantified. A reporter gene consists of a regulatory DNA sequence (a *cis*-regulatory element or a promoter) coupled, in the classical version, to an open reading frame of a reporter protein. The most widely used reporter proteins are: (i) fluorescent proteins, (ii) enzymes, (iii) selection markers such as antibiotic resistance proteins, proteins conferring sensitivity to a factor supplied with medium or metabolic enzymes that rescue corresponding biosynthetic pathway in a deficient strain. The last category is usually not used for quantitative analysis but rather for high throughput applications where the selection of only positive hits is desired.

In contrary, a variety of reporter enzymes are being used for highly sensitive quantitative analysis. Depending on the type of the substrate used, the enzyme activity is measured by chemiluminescence, fluorescence or light absorption. The enzymatic nature of the readout results in signal amplification as each molecule of the enzyme converts multiple substrate molecules at very high speed leading to high signal intensities. Therefore, enzymatic assays can pick up subtle signals and can be performed with low cell numbers. Among the most often used reporter enzymes are β -Galactosidase, β -Lactamase, placental alkaline phosphatase and different Luciferase enzymes.

Reporter gene assays are widely used and have become a golden standard for transcription factor activity measurement. They have been successfully used for drug discovery purposes in high throughput screening of chemical libraries. For example such assays have proven to be useful in the search for G-protein coupled receptor (GPCR) agonists, the procedure called receptor deorphanization ([Hill, Baker et al. 2001](#); [Johnston 2002](#)). A prototype assay is carried out using a cell line stably transfected with a reporter construct carrying a *cis*-regulatory element that is indicative of the receptor activation. Ligand efficacy is assessed by the reporter gene expression level. For more informative assays, several different reporter constructs were used coupled to different reporters ([Kotarsky, Nilsson et al. 2003](#)). For example the *c-fos* promoter was often used in reporter constructs since it carries several different *cis*-elements. These simple assays are used to get a first indication of a possible interesting hit. By using multiple and diverse reporters much more

information could be learned about the drug responses elicited in living cells. Diverse assay formats can be coupled to reporter gene expression that covers a considerable portion of all cellular events involved in signal transduction. For example detailed analysis of the transcriptional regulation can be done by measurement of promoter activities regulated by multiple transcription factors compared to activities of isolated *cis*-regulatory elements that are bound by specific transcription factors. For the proteins that regulate transcription but don't act as transcription factors themselves protein/DNA binding can be assessed by one hybrid assay ([Deplancke, Dupuy et al. 2004](#)). Protein-protein interactions in the nucleus are measured by two hybrid systems ([Luo, Batalao et al. 1997](#)). Regulated protein-protein interactions at the plasma membrane and in the cytosol constitute the main body of signal transduction networks and can be measured by a variety of protein complementation assays coupled to a transcriptional readout ([Stagljar, Korostensky et al. 1998](#)). Reporter gene assays are sensitive and flexible tools that allow measurement of a wide variety of different cellular events by selecting appropriate regulatory DNA sequences and assay designs.

Multiplexed reporter gene assays

Since reporter gene assays rely on a transcriptional readout, they enable the transition towards multiplexed assay formats by performing the analysis at the level of mRNA expression. In contrast to protein reporters, nucleic acid-based reporters can be reliably amplified up to high numbers. They can be also easily quantified in highly parallel manner by microarray hybridization or high throughput sequencing. One such multiplexed reporter system for quantification by using unique expressed oligonucleotide reporters was already mentioned above in the context of barcode arrays. The quantification was performed by microarray hybridization based on unique sequence of each reporter ([Li, Jiang et al. 2006](#)).

An alternative version of multiplexed reporter gene assay was developed by using essentially the same sequence with a restriction site placed at different positions along the oligonucleotide. After endonuclease digestion the identity of the reporters was determined by the size of the resulting fragments. The abundance of each reporter was revealed by the size of the corresponding peak obtained by the capillary electrophoresis ([Romanov, Medvedev et al. 2008](#)). The approach was applied for

simultaneous analysis of 50 different *cis*-regulatory elements. Such a reporter system can easily be applied and is homogenous with respect to the performance of the individual reporters, however, the number of assays is limited to several hundreds and can not be scaled up to perform comprehensive analysis at the genome wide level.

Reporter gene assays are very useful for further development of multiplexed approaches to elucidate cellular signaling because they provide large scale measurement options and allow coupling of various types of cellular assays to the same readout.

Split-TEV assay

Split-TEV is a protein complementation assay that is based on the reconstitution of tobacco etch virus NIa protease (TEV-protease) activity when two proteins of interest are found in close proximity for example during protein-protein interaction ([Wehr, Laage et al. 2006](#)). The TEV protease has high substrate specificity for a defined peptide motif (TEV-site) that is not found in mammalian proteins ([Phan, Zdanov et al. 2002](#)). In order to enable a protein-protein interaction assay, the TEV protease is split into two fragments: a N-terminal (N-TEV) and a C-terminal (C-TEV) part in a way that the catalytic residues of the active site are shared between the fragments and thus neither of the fragments alone can exhibit proteolytic activity. Fusion constructs are generated to link each of the putative interaction partners to one of the TEV-protease fragments in a single polypeptide chain. Therefore, when the proteins of interest interact, the TEV fragments are brought into close proximity and can fold together to form a functional protease. Protease reconstitution is monitored by the cleavage of a specific substrate peptide to trigger the release of a transcription factor that translocates to the nucleus and activates the transcription of a corresponding reporter gene (**Figure 1**) ([Wehr, Laage et al. 2006](#)). In the case when transmembrane proteins are studied, no additional anchoring domain is required since the TEV-site and the transcription factor can be attached directly to one of the interaction partners via N-terminal TEV fragment and hence kept at the plasma membrane.

Compared to commonly used two-hybrid systems that rely on transcription factor reconstitution, the Split-TEV approach involves more steps from the actual interaction to the readout. However this is a trade off for one of the main advantages of the system which is high contextual specificity. Namely, protein-protein interactions can

be monitored at their native cellular localization and their function assessed at their native environment.

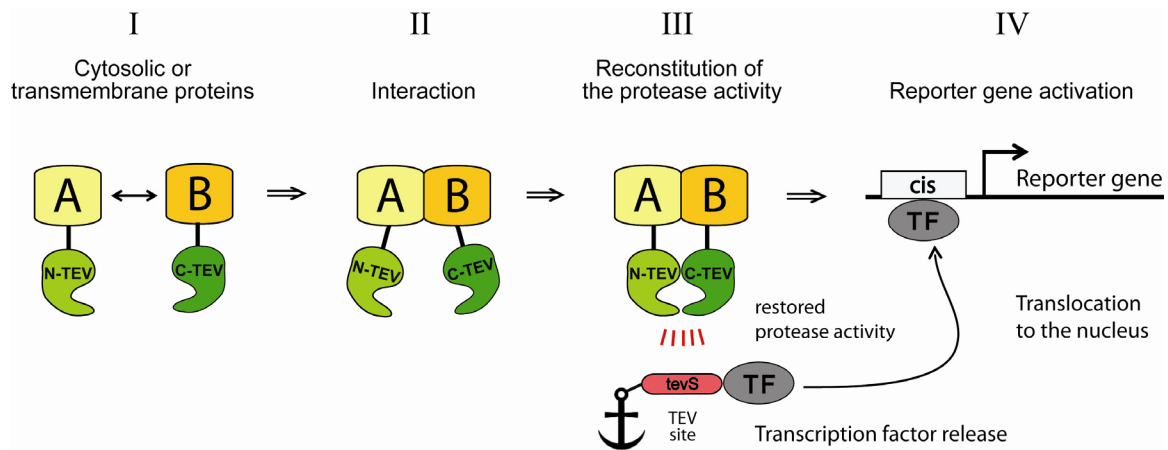


Figure 1. The principle of the split-TEV assay. Split-TEV is a proximity based protease reconstitution assay for the measurement of protein-protein interactions with high contextual specificity. Two proteins of interest (A and B) are expressed as fusion constructs with the N- and the C-terminal fragments of the tobacco etch virus protease (TEV). Upon interaction between the two proteins the inactive protease fragments are brought into close proximity and the protease activity is restored. This event is detected by the cleavage of a special proteolysis reporter protein that contains a transcription factor (TF) linked by the specific TEV protease cleavage site (*tevS*) to an anchor domain that ensures the cytosolic retention of the protein. Activation of the TEV protease leads to the release of the transcription factor that can then translocate to the nucleus and activate the transcription of an appropriate reporter gene.

Dimerization of full length receptor tyrosine kinases were measured in response to ligand binding and phosphorylation dependent interaction of cytosolic proteins Bad and 14-3-3 was analyzed in the context of signal transduction ([Wehr, Reinecke et al. 2008](#)). The ability to measure various types of interactions makes Split-TEV a well suited approach for studying signal transduction. Compared to other approaches that possess overt contextual specificity, Split-TEV does not rely on cell endogenous factors. The possibility of a transcriptional readout opens a way for multiplexed measurement of Split-TEV assays. A similar proximity based assay was developed using full length TEV protease. The TEV substrate is exposed to the protease that is recruited through protein-protein interaction leading to the transcription factor release ([Barnea, Strapps et al. 2008](#)).

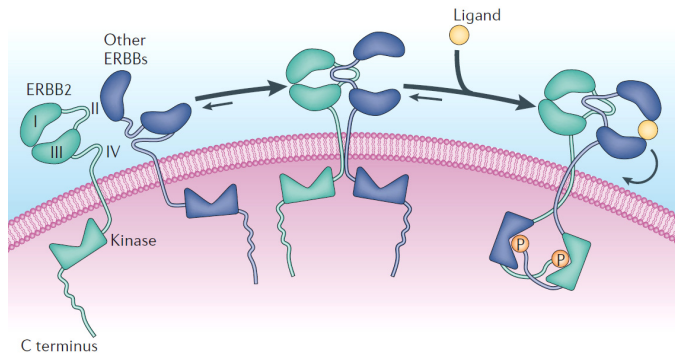
Other protein complementation assays

Among other well known protein complementation assays coupled to a transcriptional readout are split-ubiquitin, split-intein and mammalian protein-protein interaction trap assay (MAPPIT). The split-ubiquitin approach relies on endogenous proteasome

activity to release the transcription factor after ubiquitin reconstitution upon protein-protein interaction ([Stagljar, Korostensky et al. 1998](#)). The split intein system uses the mechanism of protein splicing intrinsic to the intein proteins to generate free a transcription factor upon intein reconstitution ([Kanno, Ozawa et al. 2006](#)). The MAPPIT assay is using a STAT recruitment deficient cytokine receptor that does not carry tyrosine residues fused to a protein of interest. Another protein of interest is fused to a functional phosphotyrosine domain that upon interaction rescues the signaling resulting in STAT transcription factor activation. ([Eyckerman, Lemmens et al. 2002](#)). This system is using bulky fusion proteins and depends to a large extent on endogenous signaling machinery and an endogenous transcription factor. Many more protein complementation assays (PCA) are available that, however, do not involve transcriptional readout and rely on reconstitution of a fluorescent protein (GFP variants) or an enzyme (e.g. Dihydrofolate reductase, β -Galactosidase, firefly or renilla Luciferase) reviewed in ([Michnick 2003](#)).

ErbB signaling pathway

Receptor tyrosine kinases are a class of cell surface receptors that utilize their intrinsic tyrosine kinase activity to transduce extracellular signals across the plasma membrane. A prototypic family of receptor tyrosine kinases is represented by the four closely related ErbB proteins ErbB1, ErbB2, ErbB3 and ErbB4 (alternatively named HER1, 2, 3 and 4 in humans). ErbB1 is also known as epidermal growth factor receptor (EGFR). The pathway that is initiated by the ErbB-receptor family is among the most extensive studied signaling networks and has been implicated in cancer, development and psychiatric diseases ([Citri and Yarden 2006](#); [Mei and Xiong 2008](#)). Due to their role in breast, lung and colorectal cancer ErbB proteins have become major targets for anticancer drug development ([Hynes and Lane 2005](#)). Under normal physiological conditions, ErbB receptors regulate cell proliferation, differentiation and apoptosis and play important roles in development ([Holbro and Hynes 2004](#)). ErbB receptors bind numerous growth factor ligands among which is the family of the neuregulin proteins that are encoded by 4 different genes (Nrg1-4). The best characterized is neuregulin 1 (Nrg 1).



In the inactive state, extracellular domains II and IV of ErbBs interact and mask the dimerization arm in the domain II thereby keeping the ligand binding domains I and III open. Upon ligand binding, the domains II and IV are separated and the dimerization arm is exposed triggering back-to-back dimerization of the activated receptors. In the case of the ligand binding deficient ErbB2 protein, the

Figure 2 Domain structure and dimerization mechanism of the ErbB receptors.

(adapted from [Citri & Yarden, 2006](#)).

The family consists of 31 protein isoforms produced by alternative promoter usage and alternative splicing. Some of the neuregulin isoforms are transmembrane proteins involved in cell contact signaling and others are shed and function as paracrine ligands but all of them share a conserved EGF domain that is required for signaling ([Mei and Xiong 2008](#)). All of the ErbB family members are single-transmembrane proteins that carry four domains in their extracellular region. Domains I and III contain cysteine rich clusters important for binding of peptide ligands. Domains II and IV are important for dimerization mechanism. Upon ligand binding, a dimerization arm is extended that enables homo and heterodimerization between the members of the family. In the case of ErbB2, the ligand binding domains I and III interact with each other leaving no place for the ligand, rendering the receptor deficient in ligand binding and permanently dimerization competent (**Figure 2**) ([Garrett, McKern et al. 2003](#)). On the intracellular side, the ligand binding triggers activation of the tyrosine kinase domain that leads to auto- and cross-phosphorylation of the intracellular domains of both partners. The pattern of phosphorylated tyrosine residues is thought to determine downstream signaling pathways by recruitment of about 10 different signaling proteins containing either SH3 or PTB phosphotyrosine binding domains. The repertoire of binding partners for each ErbB-family member was characterized using protein arrays as well as pulldown experiments with peptides flanking all putative phospho-tyrosine residues ([Schulze, Deng et al. 2005](#); [Jones, Gordus et al. 2006](#)).

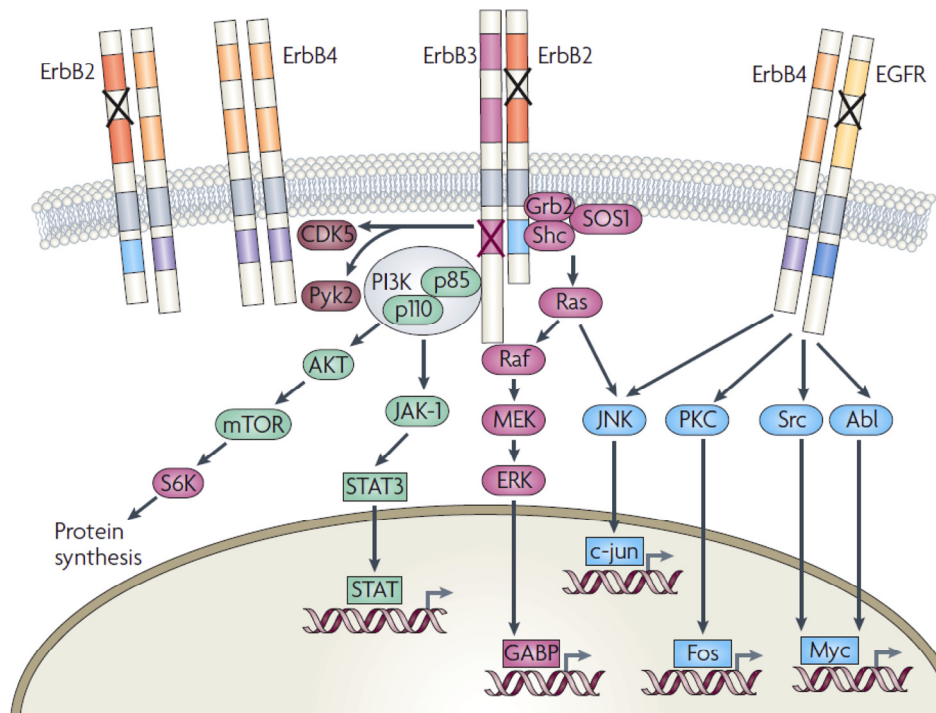


Figure 3. An overview of the Neuregulin/ErbB signaling. (adapted from [Mei & Xiong, 2008](#)). Transmembrane receptor tyrosine kinases of the ErbB family can form homo- and heterodimers upon binding of the neuregulin (Nrg). Ligand binding to ErbB3 and ErbB4 receptors initiate various downstream signaling cascades by recruitment of phosphotyrosine binding proteins such as Grb2, SHC, PI3K, etc. leading finally to activation of certain transcription factors and changes in gene expression. Two members of the ErbB family ErbB2 and ErbB3 are nonautonomous since ErbB2 can not bind ligands and ErbB3 carries an inactive tyrosine kinase domain. However, these two proteins form the most signaling potent heterodimer ([Alimandi, Romano et al. 1995](#)). This is probably due to the fact that ErbB2 is a strong dimerization partner and positive regulator for all ErbB proteins. The ErbB3 C-terminal domain exhibits enhanced activity in recruitment and activation of the downstream signaling proteins when phosphorylated by ErbB2. Only those receptor combinations are shown that are involved in Nrg signaling. ErbB2 and epidermal growth factor receptor (EGFR) do not bind Nrg which is designated by the black crosses.

All of the ErbB receptors interact with SHC (Src homology 2 containing) and Grb2 (growth factor receptor bound-2) proteins that are responsible for activation of MAP kinase pathway via SOS and Ras small GTPase (**Figure 3**). Another prominent effector protein is the p85 subunit of the PI3K (phosphatidylinositol 3 kinase) that is recruited to six phosphotyrosine sites at the ErbB3 receptor and one site at the ErbB4. Apart from that, other important downstream effectors of the ErbB signaling are: signal transducer and activator of transcription proteins (STATs), SRC tyrosine kinase, phospholipase C gamma (PLC γ), mammalian target of rapamycin (mTOR) and protein kinase C ([Olayioye 2001](#)).

It is important to point out that out of four members of the ErbB family only two (ErbB1 and ErbB4) can operate independently as homodimers because ErbB2 is

deficient in ligand binding and ErbB3 deficient in tyrosine kinase activity. However, both of them are fully functional as parts of heterodimers with other ErbB receptors (**Figure 3**). Moreover heterodimers comprised of ErbB2 and ErbB3 exhibit high signaling activity which is conferred by the strong signaling potency of the ErbB3 C-terminal tail ([Wallasch, Weiss et al. 1995](#); [Waterman, Alroy et al. 1999](#)). Furthermore, ErbB3 is thought to undergo an internalization pathway leading to its recycling back to the plasma membrane and not to the degradation as for the other ErbB proteins which may further contribute to the enhanced signaling ([Waterman, Sabanai et al. 1998](#)). On the other hand ErbB2 has been shown to act as a preferred heterodimerization partner and a positive regulator for all the other ErbB receptors ([Graus-Porta, Beerli et al. 1997](#)). Despite its inability to bind ligands, overexpression of ErbB2 can elicit signaling leading to transformation of cell lines in culture and contributing to many human malignancies in particular breast cancer ([Di Fiore, Pierce et al. 1987](#); [Olayioye 2001](#)). With multiple ligands and receptor combinations the ErbB-Nrg signaling pathway is a large and interconnected network that obviously needs to be analyzed at the systems level. Since many years, efforts are directed towards computer modeling of the ErbB signaling, however, there is still a significant lack of quantitative dynamic data that would characterize various aspects of signal transduction ([Kholodenko, Demin et al. 1999](#); [Schoeberl, Eichler-Jonsson et al. 2002](#)). Therefore, high throughput technologies are required to generate such data for different cell types.

Aim of the study

Here, we present a novel integrated and scalable reporter system - which we termed EXTassay - to simultaneously monitor cellular signaling at multiple levels within living cells and under identical experimental conditions. The system can accommodate versatile reporter gene formats. In a proof of principle experiment, we combine split-TEV and *cis*-regulatory element analysis to simultaneously monitor ErbB receptor activation at the plasma membrane and downstream signaling at the level of transcription factors. Using different split-TEV assays, we quantitatively measure neuregulin1 (Nrg1) regulated ErbB receptor dimerization and phosphorylation-dependent recruitment of first order signaling mediators ([Wehr et al 2006, 2008](#)). In parallel, we measured the activity of more than twenty different *cis*-regulatory elements to monitor the activity of different transcription factors to quantitatively assess their role in ErbB receptor mediated signal transduction.

MATERIALS

Laboratory supplies and equipment.

Equipment

Picodrop Spectrophotometer	Picodrop Limited
Agilent 2100 Bioanalyser	Agilent Biotechnologies
7500 Fast Real-time PCR System	Applied Biosystems
Microplate reader Mitras LB940	Berthold Technologies
NimbleGen 4-bay Hybridization System	Roche NimbleGen
QPix Colony Picking Robot	Genetix
Microarray Scanner G2565AA	Agilent Biotechnologies
Hybridization oven G2545A	Agilent Biotechnologies
GenePix 4200A microarray scanner	Axon Instruments
Biophotometer	Eppendorf
Heraeus Biofuge Fresco	DJB Labcare
Sorvall RC-5B Plus Ultracentrifuge	Thermo Scientific
Centrifuge 5810R	Eppendorf
Concentrator 5301	Eppendorf
Thermocycler T3	Biometra
Thermocycler TGradient	Biometra
Hera Cell 150 Cell Culture Incubator	Thermo Scientific
Hera Safe KS 12 Safe Workbench	Thermo Scientific
Schaking Incubator Multitron 2	Infors HT
Electrophoresis power supply	Pharmacia LKB
UV Gel Documentation System	iNTAS
Thermomixer comfort	Eppendorf
Arium 611 Water Purification System	Sartorius
Leica DM IRBE inverted microscope	Leica
Ultra-Low Temperature Freezer U725	New Brunswick Scientific
“GenePulser” electroporation device	BioRad

Software

Windows XP Professional	Microsoft
Microsoft Office	Microsoft
Acrobat Reader 7.0	Adobe
Endnote X2	Thomson Research Soft
Illustrator CS3	Adobe
Lasergene 7.0.0	DNA Star Inc.
Vector NTI Advance 10	Invitrogen
GenePix Pro 6.1	Axon
Feature extraction software	Agilent
MicroWin 2000	Berthold Technologies
R (statistical computing environment)	Open source

Plastic ware

General laboratory materials from Eppendorf, Falcon, Gilson and ABgene were used for molecular biology applications. For cell culture applications plastic wear from Falcon and Nunc was used.

Kits:

QIAprep Spin Miniprep Kit	Qiagen
QIAquick Gel Extraction Kit	Qiagen
QIAfilter Midi und Maxi Kit	Qiagen
NucleoSpin Plasmid Quick Pure	Macherey-Nagel
NucleoBond PC100	Macherey-Nagel
NucleoSpin Extract II	Macherey-Nagel
ChargeSwitch PCR Clean-Up Kit	Invitrogen
T7 Megascript Kit	Ambion
T3 Megascript Kit	Ambion
NimbleGen Wash Buffer kit	Roche NimbleGen
NimbleGen Sample Tracking Control Kit	Roche NimbleGen
NimbleGen Hybridization Kit	Roche NimbleGen
Pathway profiling Luciferase system 1	Clontech
Pathway profiling Luciferase system 4	Clontech
PCR Core Kit Plus	Roche
Transcriptor High fidelity cDNA Synthesis Kit	Roche
RNeasy Mini Kit	Qiagen
RNase-Free DNase Set	Qiagen
Deoxynucleoside Triphosphate Set	Roche

Reagents

General reagents

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

Agarose	Appllichem
Agar	BD
Yeast extract	BD
Bacto-Pepton	BD
Trypton	BD
GeneRuler 1 kb DNA ladder	Fermentas
GeneRuler 100 bp DNA ladder	Fermentas
6x DNA Loading Dye	Fermentas
Power SYBR Green PCR Master Mix	Applied Biosystems
Trizol reagent	Invitrogen
5-(3-aminoallyl)-UTP	Ambion
Hydroxylamine (4M)	Ambion
Coupling buffer	Ambion
Cy5 Mono-Reactive Dye Pack	GE Healthcare
Cy3 Mono-Reactive Dye Pack	GE Healthcare
Acetonitril	J.T.Baker
Fomamide	Sigma-Aldrich
SSPE buffer (20x)	AppliChem
Salmon Sperm DNA	Sigma-Aldrich
Yeast tRNA	Sigma-Aldrich
Blocking agent	Agilent Technologies
HI-RPM microarray hybridization buffer	Agilent Technologies
SSPE buffer (20x)	AppliChem

Cell culture media and reagents

Dulbecco's Modified Eagle Medium	Lonza
OptiMEM reduced serum medium	Gibco
McCoy's 5A medium	Gibco
Fetal Bovine Serum	Gibco
Horse Serum	Gibco
GlutaMax,	Gibco
Penicillin/Streptomycin	Lonza
Trypsin	Gibco
Poly-L-Lysine	Sigma-Aldrich
DMSO	Sigma-Aldrich
Laminin	Sigma-Aldrich
Lipofectamine 2000 Transfection Reagent	Invitrogen
NGF	Peptotech
Nrg1b	from Dr. T.Fischer
dbcAMP	BioLog, Bremen
Doxycyclin	Alexis Biochemicals
Rapamycin	Alexis Biochemicals

Enzymes

Restriction enzymes	New England Biolabs
REDTaq DNA polymerase	Sigma-Aldrich
HotStarTaq DNA polymerase	Qiagen
Taq polymerase	Roche
<i>Pfu</i> Ultra DNA polymerase	Stratagene
Easy-A HiFi PCR cloning enzyme	Stratagene
T4 DNA ligase	Promega
RQ1 DNase	Promega
BP clonase II	Invitrogen
LR clonase II Plus	Invitrogen
LR conase	Invitrogen
Superscript III Reverse Transcriptase	Invitrogen

Microarrays

Agilent 8x15K	Agilent Biotechnologies
NimbleGen 4x72K	Roche NimbleGen
CodeLink Activated Slides	Amersham Biosciences

Molecular biology solutions

TAE 50x

Tris-Acetate, pH 8.0	2 M
EDTA	50 mM
Glacial acetic acid	57,1 ml
adjust the volume to 1 L with H ₂ O	

TE (Tris-EDTA)

Tris-Cl	10 mM
EDTA	1 mM
adjust the pH to 7,4	

PEG solution

PEG 8000	30%
MgCl ₂	30 mM

Ethidiumbromide

EtBr 1% in H ₂ O	10 mg/ml
Final concentration in a gel	1 µg/ml

DNA loading buffer

Bromphenolblue	0,25%
Xylencyanol	0,25%
Ficoll (Typ 400)	15%

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)

LB-medium (Luria and Bertani medium)

Yeast extract	0,5%	(w/v)
Bacto-Pepton pH7,5	1%	(w/v)
NaCl	1%	(w/v)

TB-medium (Terrific broth)

Trypton	12 g
Yeast extract	24 g
Glycerin	4 ml
H ₂ O	900 ml
Autoclave and cool down to 60°C	
add 20 ml of autoclaved 50x phosphate buffer:	
KH ₂ PO ₄	0,17 M
K ₂ HPO ₄	0,72 M

Bacterial stock freezing medium

Glycerol	65% (v/v)
MgSO ₄	0,1 M
Tris-Cl, pH8	25 mM
The solution has to be autoclaved	

SOC-Medium

Bacto-Pepton	2%	(w/v)
Yeast Extract	0,5%	(w/v)
Glucose	20 mM	
NaCl	10 mM	
KCl	2,5 mM	
MgSO ₄	10 mM	

Antibiotics

	f. c. in LB medium or agar plates
Kanamycin	50 µg/ml
Ampicillin	100 µg/ml
Carbenicillin	100µg/ml
Gentamycin	25-50 µg/ml

LB-Agar plates

Yeast extract	0,5%	(w/v)
Bacto-Pepton pH7,5	1%	(w/v)
NaCl	1%	(w/v)
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 μ g/ml
IPTG (Isopropyl- β -D-thiogalactopyranosid)	15 μ g/ml

Microarray related buffers and solutions**MES 12x**

MES hydrate	64,61 g
MES Sodium salt (1,22 M MES, 0,89 M Na ⁺) pH 6,5-7,5	139,3 g

2 x Affymetrix Hybridization Buffer

MES	10 mM
NaCl	1M
EDTA	20 mM
Tween-20	0,01%

Coupling buffer for the aaRNA labeling

Na ₂ CO ₃ , pH 9,0	0,3 mM
--	--------

Coupling Neutralization Solution

Hydroxylamine (Ambion)	4M
------------------------	----

Modified 2xHybridization Buffer (for Agilent microarrays)

Salmon sperm DNA (10mg/ml)	1 mg/ml
Yeats tRNA (4mg/ml)	0,4 mg/ml
SSC (20x)	6x
SDS (5%)	0,6%

SSPE 20x

NaCl	3,0 M
NaH ₂ PO ₄	200 mM
EDTA	20 mM
Adjust the pH to 7.4 with NaOH	

Agilent Array Wash Buffer I

SSPE	6x
N-Lauroylsarcosine	0.005%
Triton X-100	0,025%

Agilent Array Wash Buffer II

SSPE	0.06x
N-Lauroylsarcosine	0.005 %
Triton X-100	0,025%

Luciferase assay buffers

Passive Lysis buffer

5x Passive lysis bufer from Promega diluted in H₂O

Firefly Luciferase Assay buffer

Tricine	20 mM
(MgCO ₃) ₄ *Mg(OH) ₂ *5H ₂ O	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

Renella 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

KxPO₄ (pH 5.1)

Prepare 1M KH₂PO₄ solution, adjust the pH to 5,1 using 2 M KOH

Solutions for cell culture

PBS 10x (Phosphate buffered saline)

NaCl	1,7 M
KCl	34 mM
Na ₂ HPO ₄ *2H ₂ O	40 mM
KH ₂ PO ₄	18 mM
Adjust the pH to 7,2 with 1N NaOH	

2x Freezing medium for eukaryotic cell lines

DMEM	40%
DMSO	20%
FBS	40%

PLL 250x

Poly L-Lysin in H ₂ O	5 mg/ml
Final concentration	0,02 mg/ml

Laminin

Laminin stock solution	1 mg/ml
has to be thawed slowly at 2-8°C and diluted in PBS	
Final concentration	1 µg/ml

Penicillin-Streptomycin

Potassium Penicillin	10000 U/ml
Streptomycin Sulfate	10000 µg/ml

Trypsin-EDTA 10x

Trypsin	5 g
EDTA	6,84 mM

Mammalian Cell Lines

PC12	Rat adrenal pheochromocytoma cell line (ATCC) (Greene and Tischler 1976)
PC12 tet OFF	PC12 cell line stably expressing tetracycline-controlled transactivator (tTA) under neomycine resistance. (Clontech) (Greene and Tischler 1976)
HEK 293T	Human embrionic Kidnej immortalized cell line (ATCC) (DuBridge, Tang et al. 1987) also designated (HEK 293tsA1609neo)
U2OS	Human osteosarcoma cell line (ATCC) (Ponten and Saksela 1967)

(ATCC) – American tissue culture collection

Bacterial Strains

Escherichia coli transformation competent bacteria

XL1-blue Competent cells	Stratagene
ElectroMAX DH10B	Invitrogen,
One Shot Mach1-T1	Invitrogen

List of Vectors and Plasmids

Construct	Resistance
pGemT	AmpR
pDONR P1 P4	KanR
pDONR P4 P3	KanR
pDONR P3 P2	KanR
pDEST GL3-basic	AmpR
pDEST_X-ASPlink-TEV1-118-tevS-GV	AmpR
pDEST_X-ASPlink-TEV119-221-HA	KanR
pEYFPnuc	KanR
phRLuc/CMV	AmpR
pRLuc/SV40	AmpR
phRLuc/TK	AmpR
pcDNA3	AmpR
G5_TATA_Luciferase	AmpR
tetO_CMV_Luciferase	AmpR
pCMV_ERT2_tevS_GV_tevS_ERT2	KanR
pBK_Gal4-VP16	KanR
pEXPR_3xFlag_hFKBP_TEV1-118	AmpR
pQCXIX TM-FRB-N-TEV-GV	AmpR
pEXPR_mErbB2_ASPlink_TEV1-118_tevS_GV	AmpR
pEXPR_mErbB2_ASPlink_TEV119-221-HA	KanR
pEXPR_mErbB2_Flag	KanR
pEXPR_mErbB3_ASPlink_TEV119-221-HA	KanR
pEXPR_mErbB3_Flag	KanR
pEXPR_mErbB4-CYT1_ASPlink_TEV119-221-HA	KanR
pEXPR_mErbB4-CYT1_Flag	KanR
pEXPR_mShc_ASPlink_TEV119-221-HA	KanR
pEXPR_mGrb2_ASPlink_TEV119-221-HA	KanR
pEXPR_PI3K-p85 α _ASPlink_TEV119-221-HA	KanR
pEXPR_PI3K-p85 β _ASPlink_TEV119-221-HA	KanR
pENTR_L1_bGHpA-pause_L4	KanR
pENTR_L1_G5_L4	KanR
pENTR_L4_G5_TATA_L3	KanR
pENTR_L4_TKprom_L3	KanR
pENTR_L4_S1-TATA_Intron_L3	KanR
pENTR_L4_S2-TATA_Intron_L3	KanR
pGL3_G5_TATA-Intron_EXT_Luci	AmpR
pGL3_G5_TKprom_EXT library_Luci	AmpR
pGL3_bGHpA-pause_Cis-TATA_EXT library_Luci	AmpR
pGL3_bGHpA-pause_G5-TATA_EXT library_Luci	AmpR
pENTR_L3_EXT library_L4	KanR

List of Primers

Cloning primers

10243	Dec1_s	AGCTAGTTGCTAAGTCTGCCGAGTAG
9752	Dec2_as	TCGTACATGCATTGACTCGCGTCTAC
7503	Spot_amine_KS_s	GAGGTCGACGGTATCTAGGTGACACTAT
7505	Spot_SK_as	CTCTAGAAGTCTAGTGGATCGACTCACTATAGG
10244	Dec1_T3_s	CTCGAGAGCTAGTTGCTAAGTCTGCCGAGTAGAATT AACCCCTCACTAAAGGGTAGGTGACACTAT
9754	Dec2_T7_as	CTCGAGACTAGTTCGTACATGCATTGACTCGCG TCTACTAATACGACTCACTATAGG
7692	B1_mGRB2_s	GGGGACAAGTTTGTACAAAAAAGCAGGCTTGACCATG GAAGCCATCGCCAAATATGACTTCAAAG
7693	B1_mGRB2_s	GGGGACAAGTTTGTACAAAAAAGCAGGCTTGAA GCCATCGCCAAATATGACTTCAAAGC
7694	B2_mGRB2_as	GGGGACCACTTTGTACAAGAAAGCTGGGTCGA CGTTCCGGTTCCTACTGGGGTG
7695	B2_STOP_mGRB2_as	GGGGACCACTTTGTACAAGAAAGCTGGGTCTCAG ACGTTCCGGTTCCTACTGGGGTG
7725	B1_ACCATG_mSHC_s	GGGGACAAGTTTGTACAAAAAAGCAGGCTTGAC CATGAACAAGCTGAGTGGAGGCGGC
7728	B2_Stop_mSHC_as	GGGGACCACTTTGTACAAGAAAGCTGGGTCTCAC ACTTTCCGATCCACGGGTTGC
8471	ErbB3_XhoI_as	GTTCCCTCGAGGTAATACATACCCTTGG
8470	ErbB3_B1_accATG_s	ACAAGTTTGTACAAAAAAGCAGGCTTACCATGAG TGCGATTGGGACTCTGCAGG
8472	ErbB3_XhoI_s	TTACCTCGAGGAACACAGCATGGTG
8473	ErbB3_SacII_B2_as	GAGACCGCGACTAGTGGGACCACTTTGTACAAGAA AGCTGGGTCAATTCTCTGGGCGTTAGCCTTAGGG
8551	ErbB3_ClaI_s	AATATCGATGGGTTCTGTAAGTGTACCAAG
10291	B1_G5_s	GGGGCAAGTTTGTACAAAAAAGCAGATGCCTGCAGGT
10293	B4_G5_as	CCCCCAACTTTGTATAGAAAAGTTGATACCCTCTAGAGT
12351	B1_bGHPA-pause_s	TTACAAGTTTGTACAAAAAAGCAGGCTCTAATTCGCTAGAGGGCCC
12352	B4_bGHPA-pause_as	TTCAACTTTGTATAGAAAAGTTGTTTCGATAGAGAAATGTTCTGG
14079	B4r_G5_s	GGGGCAACTTTTCTATACAAAAGTTG ATGCCTGCAGGT
14080	B3r_G5_TATA_as	GGGGCAACTTTATTATACAAAAGTTGGAGCTCGGTACC
14081	B4r_CMV prom_s	GGGGCAACTTTTCTATACAAAAGTTGCAATATTGGCCATTAGCC
14082	B3r_CMV prom_as	GGGGCAACTTTATTATACAAAAGTTGCACTGACTGCGTTAGC
10234	B4r_TK prom_s	GAGACAACCTTTTCTATACAAAAGTTGAAATGAGTCTTCGGACCTCGCGG
10235	B3r_TK prom_as	TCTCCAACCTTTATTATACAAAAGTTGTTAAGCGGGTCGCTGCAGGGTC
13529	B4r_S1_bGl-Intr_s	GGGGCAACTTTTCTATACAAAAGTTGTCCAAACGCAAGA TCCATGTGGAGAACTTCAGGGTGAGTTGGG

13530	B4r_S2_bGl-Intr_s	GGGGCAACTTTTCTATACAAAAGTTGAGGTTTGAGTCCAGC TTCCATGTGAGAACTTCAGGGTGAGTTTGGG
13531	B3r_bGl-Intr_as	CAACTTTATTATACAAAAGTTGGTTGCCCAGGAGCTGTAGGAAAAAG
11057	B4r_Stratagen_Luc_s	GGGGCAACTTTTCTATACAAAAGTTGCCACACCTCCCCCTGAACC
11053	B3r_Stratagen_Luc_as	GGGGCAACTTTATTATACAAAAGTTGCAGTACCGGAATGCCAAGCTGG
11058	B4r_pGL2-basic_s	GGGGCAACTTTTCTATACAAAAGTTGCGCCAGCCCAAGCTACCATG
11054	B3r_pGL2-basic_as	GGGGCAACTTTATTATACAAAAGTTGCCAACAGTACCGGAATGCCAAGC
11059	B4r_Clont pTA-Luc_s	GGGGCAACTTTTCTATACAAAAGTTGAGGTTACTTGAGCGGCCGC
11055	B3r_Clont pTA-Luc_as	CAACTTTATTATACAAAAGTTGGGTGGCTTTACCAACAGTACCGGAA
11060	B4r_TransLucent_s	GGGGCAACTTTTCTATACAAAAGTTGGGCTGCGCAACTGTTGGGAAG
11056	B3r_TransLucent_as	GGGGCAACTTTATTATACAAAAGTTGGGTGGCTTTACCAACAGTACCGG
16014	B4r_Dec1_EXT_s	ggggCAACTTTTCTATACAAAAGTTGAGCTAGTTGCTAAGTCTGCCGAGTAG
16015	B3r_Dec2_EXT_as	ggggCAACTTTATTATACAAAAGTTGTCGTACATGCATTGACTCGCGTCTAC
10242	B3_Dec1_EXT_s	ggggCAACTTTGTATAATAAAGTTGAGCTAGTTGCTAAGTCTGCCGAGTAG
9789	B2_Dec2_EXT_as	ggggCCTTTGTACAAGAAAGCTG TCGTACATGCATTGACTCGCGTCTAC

Sequencing primers

8474	ErbB3_sequencing	CTTGAGCCACTTAATTACCTGCCTGG
9199	ErbB3_seq1	GGTCGCTGCTTCGGGCC
9200	ErbB3_seq2	CCACCACTCTCTGAACTGGACC
10980	RVprimer3	CTAGCAAAATAGGCTGTCCC
10764	Luci_as	GGCGTCTCCATGGTGGCTTTACC
10977	pENTR_s	CGCGTTAACGCTAGCATGGATCTC
10978	pENTR_as	GACACGGGCCAGAGCTGCAG

Real time PCR primers

14032	Luci_qPCR_s	GTCTTCCCGACGATGACG
14033	Luci_qPCR_as	GTCTTTCCGTGCTCCAAAAC
14982	GL3_ff-Luci qPCR_s	GAGATACGCCCTGGTTCCTG
14983	GL3_ff-Luci_as	CCACCTCGATATGTGCATCTGT
10243	Dec1_s	AGCTAGTTGCTAAGTCTGCCGAGTAG
9752	Dec2_as	TCGTACATGCATTGACTCGCGTCTAC
9146	b-actin_s	CTTCTCCCTGGAGAAGAGC
9147	b-actin_as	ATGCCACAGGATTCCATACC
9578	Anchored oligo dT-Mix	TTTTTTTTTTTTTTTTTTTTTTTTVN
4542	Random nanomer	NNNNNNNNN
3213	B1_s	GGGGACAAGTTTGTACAAAAAAGCAGGCTCT
13446	B2_as	ACCACTTTGTACAAGAAAGCTG
116201	SP6_TK_s	gagagagaTTTAGGTGACACTATAGAATCCTGCAGCGACCCGCTTAA
116202	pA_Luci_as	TTTTTTTTTTTTTTTTTTTTTTTTTTTTCCTGCATTCTAGTTGTGG

METHODS

Molecular biology methods.

Molecular cloning

PCR products amplified with Taq polymerase were subcloned into pGem-T vector via so called TA-cloning. The approach is based on the property of the Taq polymerase to add an overhang nucleotide (preferentially adenosine) to the 3'-end of the completed PCR product. pGemT vector is provided in a linearized form with a 5' thymidine overhang which binds complementary to the PCR product. Thereby an efficient ligation, catalyzed by the T4 DNA ligase (Promega) is enabled. DNA fragments were further excised with the help of restriction enzymes: Type II DNA endonucleases (NEB). The vector of interest was digested with the same or compatible enzymes as the insert fragment. Restriction digestion was performed with 3-5 µg of the plasmid DNA, 10-20 units of the enzyme at 37°C in the volume of 20 to 100 µl. Suitable 10x reaction buffer was selected according to the supplier's recommendations. DNA fragments were separated by electrophoresis in 1% ethidium bromide containing agarose gel in 1x TAE buffer and visualized with UV light. In order to extract selected DNA fragments, bands of the appropriate size were excised from the gel by scalpel and processed with the help of the Gel Extraction Kit (Macherey-Nagel) according to manufacturer's protocol. Purified DNA fragments (the insert and the vector backbone in molar ratio of 4:1) were combined with 1 µl of the 10x ligation buffer in 10 µl final volume and ligated with the help of T4 DNA ligase (Promega) overnight at 4°C or for 1 hour at room temperature. Ligation mix was transformed into *E.coli* strain XL-1 blue or DH-10b.

Gateway Recombination-mediated cloning

Most of the constructs were cloned with the help of recombination-based cloning systems: The conventional cloning procedure relies on type II DNA endonucleases and T4 DNA ligase to specifically cut and join the DNA fragments. It is performed in many steps and suffers problems due to suboptimal efficiency of each of those steps. In contrary, the Gateway system makes use of site-directed recombination enzymes derived from the bacteriophage lambda that are able to insert a desired DNA fragment into a construct with much higher efficiency ([Hartley, Temple et al. 2000](#)). Detailed description of the system can be found online in the product manual book (Invitrogen). In order to clone a desired DNA sequence it was PCR amplified using primers that carry specific recombination sites: *attB1* and *attB2*. The PCR product was recombined with the pDONR vector carrying *attP1* and *attP2* sites in a so called "BP reaction" catalyzed by the Gateway® BP Clonase™ II enzyme mix, containing the viral recombination protein Integrase, the *E. coli*-encoded protein Integration Host Factor and the reaction buffer. Each BP reaction included:

<i>attB</i> -PCR product	3 µl	(unpurified, 10-100 ng)
pDONR207 (GentaR)	1 µl	(100 ng)
<u>BP clonase II</u>	<u>1 µl</u>	
	5 µl	

The reaction was incubated at room temperature overnight and 1µl was used for the electroporation of the DH-10b bacteria. Positive clones were selected by Gentamycin resistance and sequenced. Bacteria transformed with unrecombined donor vector did

not produce any colonies, because the vector carried a toxic *ccdB* gene flanked by the recombination sites. During recombination, the *ccdB* gene was replaced by the respective insert. Thereby a selective growth of only the properly recombined clones was ensured. Constructs carrying the pDONR backbone and the insert are designated “entry clones” (pENTR) and serve as shuttle constructs for quick transfer of the insert into any expression vector of choice. As a result of recombination between the *attB* and *attP* sites an *attL* site is produced. To generate the final expression construct (pEXPR) the *attL* sites of an entry clone were recombined with the *attR* sites of the so-called destination vector (pDEST). This recombination termed “LR reaction” is catalyzed by the Gateway® LR Clonase™ II enzyme mix containing the viral proteins Integrase and Excisionase, the Integration Host Factor and the reaction buffer.

Each LR reaction contained:

pENTR (25-100 ng/μl)	1 μl
pDEST (100 ng/μl)	1 μl
LR clonase	1 μl
H ₂ O	2 μl

The reactions were incubated overnight at room temperature and transformed by electroporation into the DH-10b competent cells.

The following sequences were added to the 5'- end of the primers to provide the PCR product with the *attB1* and *attB2* recombination sites:

attB1 5'- ggggACAAGTTTGTACAAAAAAGCAGGCTCT - insert specific sequence - 3'
attB2 5'- ggggACCACTTTGTACAAGAAAGCTGGGTC - insert specific sequence - 3'

Multisite Gateway Recombination mediated cloning

Recombination-mediated cloning allows assembling of multiple fragments into one construct when more than two specific recombination sites are used (Cheo, Titus et al. 2004). We have applied the MultiSite Gateway® Pro recombination cloning system from Invitrogen to generate reporter constructs consisting of three building blocks: the *cis*-regulatory element, the minimal promoter and the EXT reporter. Detailed description of the system can be found on the official web page of Invitrogen. The principle of the three-fragment recombination is shown in **Figure 4**. It relies on the high specificity with which each *att*-site is recombined only with its respective counterpart. In contrast to the single fragment Gateway recombination, in the MultiSite recombination system not all entry clones carry *attL* recombination sites. The entry clone of the middle fragment carries *attR* sites and recombines with the *attL* containing entry clones of the flanking fragments that at the same time recombined with the destination vector. As a result all three fragments are fused together and inserted into the vector in the course of one LR reaction. Fragments of interest were PCR amplified with primers introducing appropriate recombination sites, depending on the desired position of the fragment:

attB1 5'- ggggACAAGTTTGTACAAAAAAGCAGGCTCT - insert specific sequence - 3'
attB4 5'- ggggCAACTTTGTATAGAAAAGTTG - insert specific sequence - 3'

attB4r 5'- ggggCAACTTTTCTATACAAAGTTG - insert specific sequence - 3'
attB3r 5'- ggggCAACTTTATTATACAAAGTTG - insert specific sequence - 3'

attB3 5'- ggggCAACTTTGTATAATAAAGTTG - insert specific sequence - 3'
attB2 5'- ggggACCACTTTGTACAAGAAAGCTGGGTC - insert specific sequence - 3'

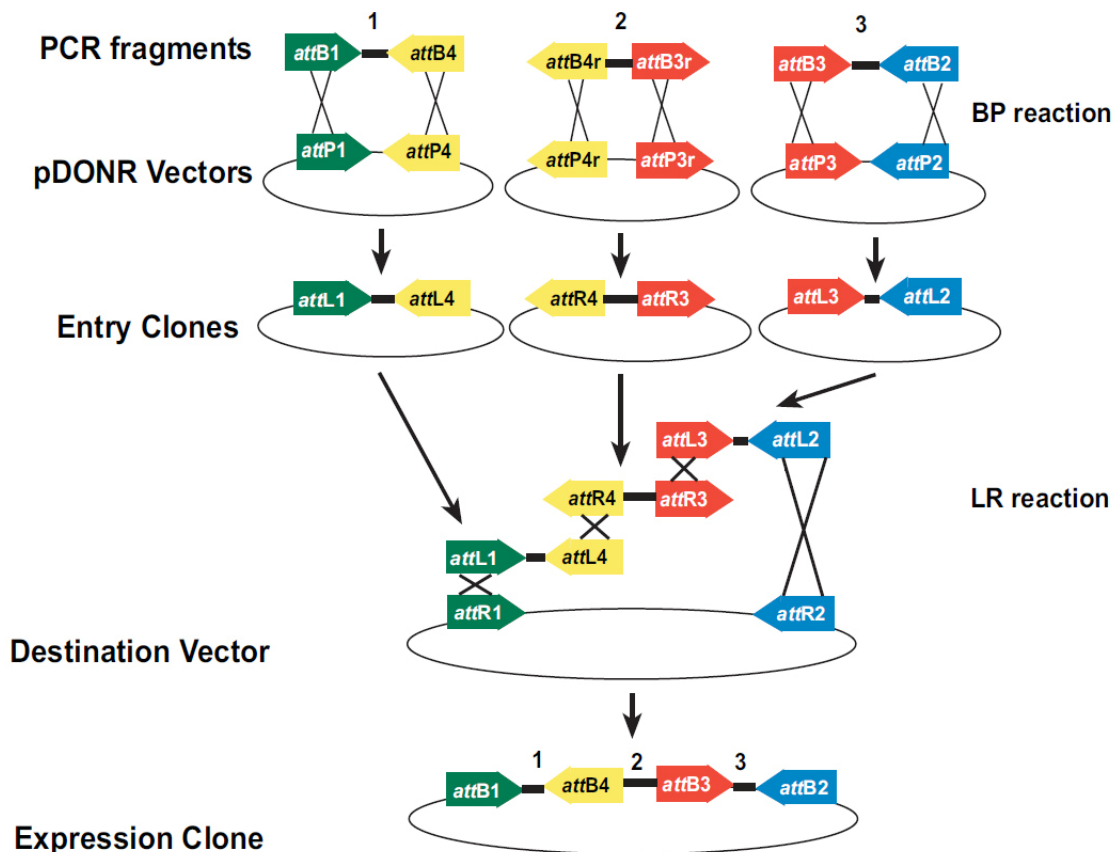


Figure 4 Schematic representation of the three-fragment recombination cloning. The image is adopted from the Multisite Gateway[®] Pro user manual and shows how three fragments can be cloned simultaneously into one construct using multiple specific recombination sites.

The PCR products were recombined with the corresponding pDONR vectors in a BP reaction identical to that of the single fragment gateway:

PCR product	3 μ l	(unpurified, 10-100 ng)
pDONR (KanR)	1 μ l	(100 ng)
BP clonase II	1 μ l	
	5 μ l	

PCR product (B1-B4)	–	pDONR (P1-P4)	–	pENRT (L1-L4)
---------------------	---	---------------	---	---------------

PCR product (B4r-B3r)	–	pDONR (P4r-P3r)	–	pENTR (R4-R3)
-----------------------	---	-----------------	---	---------------

PCR product (B3-B2)	–	pDONR (P3-P2)	–	pENTR (L3-L2)
---------------------	---	---------------	---	---------------

The reactions were incubated overnight at room temperature and transformed into *E.coli* strain DH-10b. Positive clones were selected by kanamycin resistance. Plasmid DNA was purified, tested by restriction digest and the inserts were sequenced.

Multisite LR reactions as well as all the other recombination reactions were performed in reduced volume relative to that recommended by the manufacturer:

pENRT (L1-L4)	1 μ l	5 fmol	
pENTR (R4-R3)	1 μ l	5 fmol	
pENTR (L3-L2)	1 μ l	5 fmol	or 6 fmol (3 x 2 fmol)
pDEST (R1-R2)	1 μ l	10 fmol	
LR clonase II Plus	1 μ l		

Prior to recombination the plasmid DNA concentration was measured. Entry clones were diluted to the concentration of 5 fmol/ μ l and the destination vector to 10 fmol/ μ l as higher concentrations would reduce the recombination efficiency. The following formula was used to calculate the corresponding amount of DNA in nanograms:

Where N is the size of the DNA in base pairs.

$ng = (fmol)(N)\left(\frac{660fg}{fmol}\right)\left(\frac{1ng}{10^6fg}\right)$ The LR reaction products were transformed into DH-10b or Mach1 *E.coli* strains which were then grown at 30°C to protect the insert integrity. Propagating this constructs at 37°C or using more recombination permissive *E.coli* strains (e.g.XL-1blue) led to significant reduction in cloning success rate.

PEG-mediated DNA precipitation

This method is based on the size-exclusion precipitation due to the polymeric properties of the polyethylene glycole (PEG). It was used for the purification of PCR products larger than 300bp prior to recombination cloning or purification of the recombination reaction products before electroporation. The samples were diluted to 100 μ l with 10 mM Tris (pH 7,5-8) and 1/2V (50 μ l) of the 30% PEG-8000/30 mM MgCl₂ solution was added. The tubes were vortexed and centrifuged for 15 min at RT and 13000 rpm. The supernatant was removed and the DNA was resuspended in TE or water.

Determination of nucleic acids concentration

Concentration and purity of a nucleic acids solution can be determined by spectrophotometry. According to the Lambert-Beer law absorbance of the sample (A) is a product of concentration (C), optic path length (l) and a molar absorptivity of the substance, so called extinction coefficient (ϵ): $A=Cle$.

An absorbance of a nucleic acid depends on its composition and is a combination of the single base absorbencies. It is commonly accepted to use the following relation of optical density to nucleic acid concentration as a valid estimate:

Double stranded DNA: $1A_{260} = 50 \mu\text{g/ml}$

Single stranded DNA: $1A_{260} = 33 \mu\text{g/ml}$

Single stranded RNA: $1A_{260} = 40 \mu\text{g/ml}$

Where A_{260} is a unit of absorbance measured at 260 nm which is the wavelength of a maximum absorbance for nucleic acids. For an accurate measurement it is important that an A_{260} value falls between 0,1-1 to fit the linear range of a spectrophotometer, therefore more concentrated samples were diluted prior to the measurement.

Sample purity was assessed by measuring the absorbance at 280 and 230 nm. Possible contamination of the sample with proteins was ruled out by measuring the absorbance at 280 nm, the maximum absorbance wavelength for tyrosine and tryptophane. Absorbance at 230 nm was measured to check for the presence of guanidium salt or phenol that may be carried over from silica column or phenol extraction procedures respectively.

DNA purity criteria: Samples with A_{260}/A_{280} between 1,7 and 2 were considered to be pure.

$A_{260}/A_{280} < 1,7$ indicates protein contamination

$A_{260}/A_{280} > 2,0$ indicates RNA contamination

$A_{260}/A_{230} < 2,0$ indicates the presence of carbohydrates, salts or organic solvents

$A_{330} > 0$ indicates light scattering by particles and general turbidity of the solution.

DNA sequencing and Oligonucleotide synthesis

DNA sequencing and oligonucleotide synthesis were performed at the Institute DNA Core Facility (Department of Neurobiology, MPI for Experimental Medicine, Göttingen) under the supervision of Dr. Fritz Benseler. DNA sequences were determined by the modified dideoxy sequencing method ([Sanger, Nicklen et al. 1977](#)). The approach is based on linear DNA amplification in the presence of a single sequencing primer, regular free deoxynucleotides and modified dideoxynucleotides (chain terminators). Dideoxynucleotides stochastically incorporate into all the positions where the corresponding nucleotide is found and terminate the synthesis. A mixture of products of different length is separated by capillary electrophoresis. Each of four dideoxynucleotides carries a different fluorescent dye, whereby the products are also distinctly labeled depending on the identity of the terminal nucleotide. As the DNA fragments pass through the capillary their size and fluorescence are determined and the sequence is read.

Plasmid DNA with the concentration of 100-200 ng/μl was usually used as a template for sequencing reaction. For some applications like EXT insert screening unpurified PCR products derived from RedTaq catalyzed colony-PCR were diluted with water in a ratio of 1:10 and sequenced.

Transformation of chemically competent bacteria

Aliquots of transformation-competent bacteria (*E.coli*, strain XL-1blue or Mach1) were thawed on ice. 5-10 μl of a ligation or a recombination was added to 100 μl competent cells. The bacteria were left on ice for 30 minutes and heat shocked for 30-45 seconds at 42°C. After 2 minutes on ice 800 μl of cold LB medium without antibiotics was added. The bacteria were allowed to recover for 45 minutes at 37°C and moderate shaking. In some cases cells were incubated for 1-1,5 hours at 30°C in order to prevent plasmid recombination or loss. After the expression of an appropriate antibiotic resistance, 100μl of the bacterial suspension were plated on relevant antibiotic-containing LB-agar plate. Remaining bacteria were pelleted by centrifugation for 1 minute at 8000 rpm, resuspended in 100-200μl LB/antibiotic and plated on a second LB-agar plate. Sterile glass beads were used to evenly distribute the bacteria and the plates were incubated overnight at 37°C (or 30°C respectively). For retransformation, only 10-30 μl of the competent cells was used and only 5-10 μl out of 800 μl bacterial suspension was plated.

Electroporation of bacteria

Commercially purchased electrocompetent *E.coli*, strain DH10B were diluted 1:4 with sterile 10% Glycerol solution and aliquoted by 20ul per vial. Each aliquot was used for one electroporation or further diluted with 10% Glycerol and used for 2 or more electroporations. The bacteria were allowed to thaw on ice, 1-1,5 μl of the recombination reaction were added and the mixture was transferred into a 1 mm electroporation cuvette (BioRad) prechilled on ice. The cells were given an electric pulse of 1,8 kV, under 25 μF capacitance and 200 Ω resistance using the "GenePulser" electroporation device (BioRad). The cells were then resuspended in 800 μl LB medium without antibiotics and treated as described above for the heat shock mediated transformation.

Isolation of plasmid DNA

Plasmid DNA was prepared from fresh overnight bacterial LB cultures with OD of 2 to 3. Small scale (mini-prep) and large scale (midi-prep) preparations required 2-4 ml and 50-100 ml of the bacterial culture respectively. The preparation was done with the help of corresponding kits from Macherey-Nagel or Qiagen. Detailed description of the procedures can be found in the handbooks provided by the manufacturers. The preparation is based on the main procedures of alkaline lysis and SDS precipitation of proteins and genomic DNA ([Birnboim and Doly 1979](#)). Due to its smaller size plasmid DNA remained in the supernatant and was further purified over an anion exchange resin. The DNA was bound to the resin under appropriate high-salt conditions was washed with 80% ethanol containing buffer and eluted with 100µl 5mM or 10mM Tris/HCl pH8,5. In the case of midi-prep the DNA was further precipitated with isopropanol and resuspended in the buffer TE (10mM Tris/HCl, 1mM EDTA, pH8,5)

Generating frozen stocks of bacterial cultures

Recombinant plasmid DNA was conserved by freezing bacterial cultures, that were transformed with the respective plasmids. After transformation a single clone was selected and propagated in LB medium. Three volumes of the log-phase bacterial culture were taken per two volumes of the bacterial freezing medium (e.g. 750 and 500µl, respectively). The samples were mixed and frozen 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.

Polymerase Chain Reaction (PCR)

Polymerase Chain Reaction is a widely used method for selective *in vitro* amplification of a defined DNA sequence. During the procedure a thermostable DNA polymerase derived from the thermophilic bacterium *Thermus aquaticus* (Taq Polymerase) replicates selected template DNA molecules multiple times ([Mullis, Faloona et al. 1986](#)). Since the enzyme is unable of *de novo* DNA synthesis it requires two primer oligonucleotides that bind complementary to the outmost stretches of the template. These templates can be extended by the enzyme, to generate a full length PCR product. In each round of amplification both strands of the template are replicated since one of the primers always bind the sense and the other one to the antisense strand. In each subsequent round of amplification newly made PCR products are serving as templates leading to exponential amplification kinetics. Taq polymerase introduces about 0,8 mismatches per 1 kb per amplification cycle. For amplification of long fragments we have used PfuUltra (Stratagene), a modified *Pyrococcus furiosus* DNA polymerase which possess a 3'-5' exonuclease-dependent proofreading activity and exhibits 18-fold lower mutation rate.

Composition of the PCR reaction:

	<u>Final conc.</u>
10 x reaction buffer	
Template DNA	100 pg-25 ng/reaction
MgCl ₂ (included in the buffer or added separately)	1-4 mM
Deoxynucleotide Mix (dATP, dCTP, dGTP, dTTP)	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 ₂ O	

A PCR reaction consists of 3 main steps: melting, primer annealing and elongation which are carried out at different temperatures and are repeated multiple times (in cycles). The following rules were used for designing PCR programs:

Initial denaturation: 2-3 min at 95°C was used to remove any DNA secondary structures. This step was extended to 5-10 min when using a hot-start DNA polymerase that needs high temperatures to get activated.

Denaturation: 15-30 sec at 95°C is the first step of the cycle during which the complementary strands of the template are melted and become available for the primers.

Annealing: 20-30 sec at a temperature 2-4 degrees lower than the calculated primer T_m . In some cases an optimal annealing temperature was determined empirically by using a gradient PCR machine.

Elongation: was performed at 72°C - the maximal activity temperature for Taq polymerase. Elongation time was calculated depending on the template length and the enzyme used. DNA polymerases with proof reading activity require roughly 1 min elongation time per 1 kb template and non proofreading enzymes - 30 sec/kb.

Final extension: after desired number of cycles was completed samples were incubated another 5-10 min at 72°C to ensure that all the shorter products that might have been produced due to incomplete elongation were filled in and completed.

Designing primer sequences

Primers were designed to have a melting temperature between 50° and 60°C. However in some cases lower temperatures had to be used. Primer melting temperature can be roughly estimated by using the following formula:

$$T_m = (A + T) \cdot 2 + (G + C) \cdot 4$$

where A,T,G and C are the numbers of corresponding nucleotides in the primer sequence.

For more accurate estimation we have used an online algorithm developed by Warren A. Kibbe (Kibbe 2007) and available online at

<http://www.basic.northwestern.edu/biotools/oligocalc.html>

Colony PCR

Cloning of difficult inserts sometimes requires many clones to be analyzed before the one carrying a proper plasmid insert can be isolated. Colony PCR is a time and cost effective method for quick screening of a large number of bacterial colonies prior to plasmid DNA preparation in order to select only positive ones. The PCR is performed directly on a bacterial colony using primers specific for a positive insert or primers flanking the insert. In the later case all the inserts would be amplified and the positive ones could be distinguished by size after separation of the PCR product by the agarose gel electrophoresis. Two 96-well plates were prepared: the first one containing 20 µl per well of the complete PCR master mix and the second one containing 100 µl per well LB medium supplemented with an appropriate antibiotic. Each colony picked with a sterile tooth pick was first dipped into a distinct well on the PCR plate and then into a corresponding well on the LB plate. The PCR-plate was incubated at 95°C for 10 min. At this temperature the bacterial cells would get disrupted to release the plasmid DNA and enable amplification. After that the PCR was carried out as usual. The LB plate was sealed and kept at 4°C. After the results of the PCR were analyzed positive clones could be recovered from this plate and used for plasmid DNA preparation.

Agarose gel electrophoresis

Agarose gels of 1,5% or 3,5% were prepared by dissolving a corresponding amount of agarose in 1xTAE buffer and heating the suspension in a microwave for 5-10 min until the agarose was dissolved. The solution was cooled to roughly 60°C to reduce evaporation and ethidium bromide was added to a final concentration of 1µg/ml. The solution was poured into a gel-casting form, combs were inserted to create pockets for sample loading and the agarose was allowed to solidify for 20-30 min. The gel was then placed in a chamber 1-3 mm submerged in 1xTAE buffer. To each sample one 5th volume of the loading dye was added and the samples were loaded into the pockets of the gel. A voltage of 4-6 V/cm was applied to induce the migration of negatively charged DNA molecules towards the positive electrode leading to their size dependent separation. After desired separation was achieved the DNA fragments were visualized by the fluorescence of the associated EtBr in 260 nm UV light.

1,5% agarose gels were used for regular analysis of DNA fragments and purification of the DNA fragments for the molecular cloning. 3,5% gels were employed to visualize small differences in the PCR product length, in particular for the selection of full-length EXT clones by colony PCR.

Phenol Chloroform extraction and DNA precipitation

To purify the DNA by phenol chloroform extraction the samples were diluted to 100 µl and an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1) was added. The tubes were vortexed vigorously and centrifuged for 1 min at 13000 rpm to achieve phase separation. The upper aqueous phase was removed and mixed with 100 µl Chloroform. The vortexing and centrifugation steps were repeated. The upper phase was removed to precipitate the DNA and the lower phase was discarded.

DNA was precipitated at room temperature with ethyl alcohol in the presence of sodium acetate. Wherefore 1/10th volume of 3M NaAc pH 5,2 and 0,5 µl Glycogen was added to each DNA sample followed by vigorous vortexing. Then 2,5 volumes of 100% ethanol was added, the samples were mixed by vortexing once again and centrifuged for 15 min at 13000 rpm at room temperature. The supernatants were discarded and the pellets were washed with 70% ethanol, airdried and resuspended in 10 mM Tris pH 8,5.

Cell biology Methods

Culturing of eukaryotic cells

Eukaryotic cells were allowed to grow on 15 cm dishes at 37°C in a humidified incubator at 5% CO₂. Growth medium was supplemented with GlutaMAX (a stabilized dipeptide form of L-glutamine), penicillin, streptomycin and heat inactivated serum as described below:

PC12: DMEM low glucose medium (1 mg/ml),
1% Pen-Strep, 1% GlutaMAX,
10% FBS, 5% HS

PC12 tetOFF: same as for the PC12

U2OS: McCoy's 5A medium,
1% Pen-Strep, 1% GlutaMax
10% FBS

HEK 293T: DMEM high glucose medium (4,5 mg/ml),
1% Pen-Strep, 1% GlutaMax,
10% FBS

Poly-L-Lysin coating of plastic surfaces

For better adherence culture dishes for PC12 cells and multi-well assay plates for all the cell lines were treated with 0,02 mg/ml PLL in water. After 10 min incubation at room temperature the plates were washed twice with sterile water and dried under the laminar flow tissue culture hood. Coated plates were stored at 4°C for up to one month.

Passaging of eukaryotic cell lines

The cells were split after reaching 80-90% confluence. The growth medium was removed and the plate was washed with room temperature PBS. The cells were detached from plastic surfaces by mild treatment with Trypsin-EDTA. The reaction was terminated by adding 10 ml of growth medium containing serum and the cells were pelleted by centrifugation for 5 minutes at 800 g. The supernatant was discarded, the cells were resuspended and 1/10th of the volume was plated onto the coated dishes. PC12 cells build strong cell to cell contacts that could not be disrupted as efficiently as contacts to plastic. Therefore prior to plating, they were additionally triturated 5 times through 24G (0,55 mm) needle using 5 ml sterile syringe.

Thawing of eukaryotic cell lines

Frozen stocks of eukaryotic cell lines were kept aliquoted in 2 ml cryovials at -196°C submerged in liquid nitrogen. An aliquot was quickly thawed at 37°C, transferred into a falcon tube containing 10 ml prewarmed growth medium and centrifuged for 5 min at 800 g. The pellet was resuspended in fresh medium and plated out onto a 15 cm culture dish.

Generating frozen stocks of eukaryotic cells

The cells were preferentially frozen at the stage below 90% confluence. After trypsinization the cells were pelleted and resuspended in a small volume of ice-cold growth medium. Cells were frozen at a density of approximately 2 million cells per 1ml. 500 μ l of the cell suspension was transferred into a cryovial containing an equal volume of the ice-cold 2x freezing medium. The samples were mixed gently by inverting the tubes and frozen at -20°C . After 2 hours the vials were transferred to -80°C and after 6-12 hours to liquid nitrogen for long term storage.

Transfection of mammalian cells

The cells were typically transfected in 96-well plates for the Luciferase assay. For the microarray experiments transfections were performed in solution or in 24-well plates. Depending on the cell line and the well size, different number of cells was plated as indicated below:

Cell line	96-well	24-well	6-well
PC-12 tet OFF	50 000	250 000	1 000 000
PC12 WT	50 000	250 000	1 000 000
U2OS	15 000	100 000	400 000
HEK	20 000	100 000	400 000

Lipofectamine 2000 Transfection Reagent (Invitrogen) was used to introduce the plasmid DNA into the mammalian cells. Most of the time the cells were transfected in 96-well flat bottom cell culture plates (Falcon). The original protocol of the manufacturer was modified to achieve optimal results. Mini- as well as midi-prep quality DNA was used for transfections. Per one well in 96-well plate 10-200 ng of plasmid DNA and 0,2 μ l of the Lipofectamine 2000 were diluted in 15 μ l OptiMEM each. The two solutions were combined resulting in 30 μ l volume and incubated for 20-30 min at room temperature. The growth medium was completely removed and the cells were covered with the 30 μ l per well OptiMEM containing the DNA-Lipofectamine complexes. After 2 hours the transfection medium was exchanged for a regular growth medium. Alternatively the growth medium was supplemented with twice the amount of serum and 30 μ l per well were added on top of the transfection medium to reach the final volume of 60 μ l and the regular final concentration of serum.

Transfection in solution

For the microarray experiments PC12 tet OFF cells were transfected in suspension. The cells were trypsinized, triturated through a 24G needle 5-10 times to remove any cell clumps and the cell number was determined. Desired number of cells was pelleted by centrifugation for 5 min at 800 g and resuspended in DMEM, 1% HS to reach a cell density of 100 000 cells per 100 μ l. Depending on the cell number the following amounts of transfection components were used:

	<u>per 100 000 cells</u>	<u>per 2 000 000 cells</u>
Lipofectamine 2000	0,4 μ l	9 μ l
Renilla Mix	0,2 μ l (40 ng)	4,5 μ l (900 ng)
plasmid DNA	100 ng	6 μ g
OptiMEM	5 μ l + 5 μ l	10 μ l + 10 μ l

The DNA and the Lipofectamine 2000 reagent were diluted in one half of the OptiMEM volume each, mixed and combined together. The Lipofectamine-DNA complexes were allowed to form for 20 min at room temperature and were then added to the cell suspension. The tubes were incubated at 37°C for 4 h without shaking. Transfected cells were centrifuged to remove the transfection mixture, resuspended in fresh media mixed and plated out according to the experimental design.

Luciferase reporter gene assay

Measurements of the firefly luciferase (ff-Luciferase) activity are often normalized to the readings of a different reporter gene, e.g. Renilla Luciferase (Renilla) expressed under a constitutively active promoter. This type of normalization helps to correct for the differences in cell numbers, transfection efficiency, RNA and protein expression and general well being of the cells from well to well. We have 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 cotransfected along with any firefly reporter gene assay. In order to monitor the transfection efficiency an equal amount of pEYFPnuc, a plasmid for CMV-driven expression of the nuclear localized Enhanced Yellow Fluorescent Protein was included. (Renilla Mix : pEYFPnuc = 1:1)

Composition of the Renilla Mix:

pRLuc/SV40	100 µg
phRLuc/TK	20 µg
phRLuc/CMV	10 µg
pEYFPnuc	130 µg
10 mM Tris pH8,5	up to 1,3 ml

Final conc. 200 ng/µl (100ng/µl Renilla Luciferase constructs, 100ng/µl EYFPnuc)

Luciferase assays were always performed in 96-well format. Per well 50-200 ng of plasmid DNA was transfected including 0,1 µl (20 ng) of the Renilla Mix and 10-50 ng per construct of other plasmids. To ensure the statistical reliability of the results each assay was performed in 4-6 replicates, usually 4-6 wells on the same plate transfected with the same DNA-Lipofectamine master mix. Depending on the design of the experiment, transfected cells were allowed different time to express the recombinant proteins and reporter genes. After the experiment was finished the medium was removed and the cells were lysed with 30 µl per well Passive Lysis Buffer (Promega). The plates were incubated for 15 min at room temperature with slight shaking (200 rpm). Lysates were assayed immediately or frozen at -20°C. Prior to the measurement the lysates were transferred into a black plastic microtiter plate to reduce the light signals leakage and cross mixing due to light reflection. The dual luciferase assay measuring the bioluminescence of both Firefly and Renilla Luciferases was carried out with the help of the Microplate reader Mitras LB940 (Berthold Technologies) and the associated software MicroWin 2000.

To each well 75 µl of the ff-Luciferase substrate was injected, the reaction was allowed to stabilize for 2 sec and the light signals were collected over the next 10 sec. Then 75 µl of the Renilla substrate was injected and after a 2 sec pause the 10 sec measurement was performed. The Firefly and the Renilla luciferase have high substrate specificity and do not cross-activate. Moreover ff-Luciferase activity is inhibited by the pH conditions of the of the Renilla substrate.

The data were exported from MicroWin2000 and analyzed with the help of Excel. The Firefly readings were divided by the corresponding Renilla readings producing

values in relative luciferase units (RLU). An average over the 6 replicates was taken and the standard deviation was calculated.

Preparation of the microarray target

RNA isolation

To prepare RNA from the cell culture the medium was removed, the cells were quickly washed with PBS and lysed by adding Trizol Reagent. In case of 96-well plates the washing step was omitted. Trizol Lyzates were transferred to 2ml eppendorf tubes and vortexed to ensure complete lysis. 1/5th of a volume of Chloroform was added. The samples were vortexed thoroughly and centrifuged for 15 min at 4°C, 12 000 g.

The following amounts of the Trizol Reagent and Chloroform were used

Plate	Area	Cell number	Trizol	Chloroform
96-well	0,3 cm ²	50 000	125 µl	25 µl
24-well	2 cm ²	250 000	500 µl	100 µl
6-well	10 cm ²	1 000 000	1 ml	200 µl

Trizol Reagent is a solution of phenol and guanidine isothiocyanate that lyses the cells by dissolving the membranes and upon addition of chloroform separates into two phases: organic and inorganic. RNA remains exclusively in the aqueous (inorganic) phase ([Chomczynski and Sacchi 1987](#)). Proteins that are contained in the organic phase and genomic DNA that accumulates at the interphase can also be recovered within the same preparation ([Chomczynski 1993](#)). RNA containing aqueous phase was supplemented with equal volume of 70% ethanol and further purified over an RNeasy column (Qiagen) including an on-column DNase treatment, according to the manufacturer's protocol. Both Trizol extraction and RNeasy column are known to yield RNA of high purity that is free of genomic DNA. However samples derived from transiently transfected cells contain considerable amount of plasmid DNA that is smaller in size and unlike genomic DNA can not be efficiently removed by these methods. Even after an on column DNase digestion we have always observed a slight plasmid DNA contamination in our RNA samples detected by PCR at about 40 cycles. To minimize this contamination purified RNA was additionally treated with 3 units of RQ1 DNase in solution for 30 min and repurified over an RNeasy column.

RNA precipitation

The RNA was precipitated with ethyl alcohol and ammonium acetate salt. To 100 µl RNA 50 µl of 7,5 M NH₄Ac and 0,5 µl of Glycogen solution (20 mg/ml) was added. The RNA was then precipitated by adding 450 µl of the 100% EtOH. The samples were mixed thoroughly and centrifuged at room temperature for 30 min at 12000g.

Glycogen precipitates at the same conditions as nucleic acids and serves as a carrier that prevents RNA loss during precipitation. Presence of the carrier makes the pellet visible, the precipitation procedure more reliable and does not compromise downstream enzymatic reactions such as reverse transcription, PCR or sequencing. The RNA was washed with 100 µl of 70% EtOH, airdried and resuspended in 4-6 µl of H₂O.

First Strand cDNA synthesis

First strand cDNA was generated from 0,5 – 1 µg total RNA using Superscript III reverse transcriptase (Invitrogen). Equal amounts of RNA for each sample were precipitated and resuspended in 4,5 µl RNase free water, 1 µl (120 pmol) of the random nanomer primer (#4542) was added to the final volume of 5,5 µl. The samples were denatured for 2 minutes at 70°C and placed on ice. Then other components were added:

5x First Strand buffer	2 µl
0,1 M DTT	1 µl
dNTP mix (10mM each)	0,5 µl
Superscript III reverse transcriptase (200 U/µl)	<u>1 µl</u>
<i>Final volume</i>	10 µl

The samples were incubated for 10 min at 25°C to allow for the annealing of random nanomer primers, then for 1 hour at 50°C for the RNA dependent DNA synthesis. After that the enzyme was heat inactivated for 5 min at 85°C. The cDNA was diluted 1:8 with water and used for PCR amplification.

To monitor the cDNA quality with respect to plasmid DNA contamination for each sample a negative control was performed where all the reaction components were included except the reverse transcriptase (abbreviated as –RT control).

Quantitative real-time PCR

Real time PCR was used to determine the relative expression level of an EXT reporter for the experiment where only one EXT reporter construct was used or to monitor the amplification kinetics and assess the sample quality of an EXT reporter mixture that was further used for a microarray hybridization. Quantification of the PCR product was performed using Power SYBR Green PCR Master Mix (Applied Biosystems). SYBR Green intercalating cyanine dye contained in the master mix binds double stranded PCR products as they are being generated and makes them detectable by fluorescence. The reaction mix was assembled as follows:

cDNA (1:8 diluted)	4 µl
SYBR Green PCR Master Mix	6 µl
Forward primer (50pmol/ µl)	0,1 µl
Reverse primer (50pmol/ µl)	0,1 µl

The annealing temperature for the Dec1/Dec2 primer pair was optimized to 60°C by gradient PCR. The following amplification program was used: 95°C – 10min, cycle: 95°C – 30 sec, 60°C – 30 sec, 72° - 31 sec (fluorescence reading step), 45 cycles.

For the preparative PCR the amplification was stopped shortly before the last sample had reached saturation (at about 35-40 cycles).

In Vitro Transcription (IVT)

The IVT was performed with the help of the T7 MEGAscript Kit (Ambion) that makes use of the bacteriophage T7 RNA polymerase to transcribe the DNA downstream of the T7 promoter. EXT PCR products amplified with the Dec1 (#10243) and Dec2 (#9753) primers carry T7 promoter just before the Dec2 binding site. Unpurified PCR product amplified with the RedTaq DNA Polymerase (Sigma Aldrich) or SYBRGreen PCR Mix (Applied Biosystems) was used as a template for the in vitro transcription. In the course of this reaction 5-(3-aminoallyl)-modified UTP

was incorporated into the newly synthesized RNA which was required for its fluorescent labeling for the microarray hybridization.

The IVT reaction volume was downscaled relative to the manufacturer's protocol:

PCR product	4µl
10 x Reaction Buffer	1 µl
ATP (7,5 mM)	1 µl
CTP (7,5 mM)	1 µl
GTP (7,5 mM)	1 µl
UTP (7,5 mM)	0,5 µl
5-(3-aminoallyl-UTP) (50 mM)	0,75 µl
T7 RNA Polymerase	1 µl

The reactions were incubated overnight at 37°C. Heated lid at 42°C was used to prevent evaporation which significantly increased the IVT yield. After that, the template DNA was degraded by adding 1 µl of Turbo DNase to each sample and incubating at 37°C for 15 min. Typical yield of the EXT-IVT comprised 5-10 µg of short aminoallyl-modified RNA 111 bases in length. The RNA was purified over an RNeasy column (Qiagen) using a modified protocol described below.

Purification of short RNA

The IVT product was purified via an RNeasy mini Kit (Qiagen). In order to ensure the efficient recovery of small RNAs the manufacturer's protocol was modified. The volume of all samples was adjusted to 100 µl with RNase free water. Per each sample 100 µl buffer RLT was added and the samples were mixed. Then 400 µl of isopropanol was added, the samples were quickly mixed by vortexing, immediately transferred to the RNeasy columns and centrifuged for 15 sec at 10 000 g. The membranes were washed twice with 500 µl RPE buffer and dried by centrifugation for 2 min. The RNA was eluted in two steps with 50 µl H₂O each time to maximize the RNA yield.

Labeling of the microarray target

To generate a probe for a microarray, hybridization purified RNA carrying the aminoallyl-modified nucleotides (aaRNA) was coupled to the cyanine dye esters: Cy3 or Cy5. For each labeling reaction 5-10 µg of the aaRNA was lyophilized and resuspended in 4,5 µl Coupling buffer (Ambion). One tube of the Cy3- or Cy5-monoreactive dye (Amersham) was dissolved in 55 µl DMSO and 5,5 µl was added to each sample. After a 30 min incubation at room temperature in the dark, the reaction was neutralized by the addition 2,3 µl 4M Hydroxylamine (Ambion). The tubes were incubated in the dark for another 15 minutes and the RNA was purified via the modified RNeasy protocol as described above. The dye incorporation efficiency (*E*) was calculated as follows:

$$E \text{ (pmol/}\mu\text{g)} = \frac{\text{Dye (pmol/}\mu\text{l)}}{\text{RNA Conc (}\mu\text{g/}\mu\text{l)}}$$

A successful labeling usually resulted in a dye incorporation of 50-100 pmol/µl. Such samples were used in the microarray hybridization.

Microarray hybridization and analysis

Hybridization of Agilent microarrays

The amount of the target was adjusted to the amount of the dye and not the amount of RNA in order to compensate for the slight differences in labeling efficiency from sample to sample. The target contained 10-30fmol of the Cy-dye per EXT, which depending on the labeling efficiency constituted 0,1-0,5ng RNA per EXT. We have used modified 2xHybridization Buffer containing salmon sperm DNA and yeast tRNA to increase the hybridization stringency.

The hybridization mix for the 8-plex Agilent microarray contained:

2 x Hybridization buffer	25µl
10 x Blocking agent	5 µl
Formamide (f.c. 15%)	7,5 µl
Target	12,5 µl
<i>Final volume</i>	50 µl

The samples were mixed by pipetting, centrifuged for 1 min at 13 000 rpm and 40µl was loaded per subarray. The arrays were hybridized at 65°C for 20 hours.

The arrays were opened submerged in 250ml Wash buffer-1 prewarmed to 37°C. Up to 4 array at a time were washed at RT for 1 min in 250ml Wash buffer-1 and for another 1 min in 250ml of the Wash buffer-2. The arrays were dried by dipping for 30 seconds in 250ml Acetonitril.

Hybridization of NimbleGen microarrays

The NimbleGen arrays were hybridized according to the standard protocol provided by the manufacturer however using the Cy-dye labeled aaRNA and not DNA as a target. The amount of the target for 4-plex NimbleGen microarrays was calculated in the same way as for Agilent arrays. The arrays were hybridized at 48°C (for some control experiments at 50°C). Washing procedures were carried out using the NimbleGen wash buffer Kit following the manufacturer's recommendations. The slides were dried by dipping for 30 sec in Acetonitril and scanned with the help of the GenePix 4200A microarray scanner.

Hybridization of the self spotted microarrays

Self spotted microarrays were produced at the DNA microarray facility, Medical Faculty, University of Göttingen (<http://www.microarrays.med.uni-goettingen.de>). EXT oligonucleotides were amplified with spotting primers (#7503, #7505). PCR products were purified with the ChargeSwitch PCR Clean-Up Kit (Invitrogen) and spotted on the surface of Codelink Activated Slides (Amersham) in 50mM Na₃PO₄ buffer at pH8,5. The forward primer carried an amine modification that was used for the crosslinking of the corresponding DNA strand to the surface of the slide. Therefore, the sense strand of the PCR product remained on the spot while the antisense strand was washed away. The arrays were hybridized at 50°C using Agilent hybridization chambers and gasket slides in Affymetrix hybridization buffer containing 50% Formamide. The arrays were submerged in 2x SSC, opened and then washed for 5 min in 2x SSC, 0,1% SDS, then for 1 min in 0,2% SSC and for another 1 min in 0,1% SSC. The slides were dried in the flow of nitrogen and scanned.

Microarray Scanning and Data Extraction

All arrays were scanned with 5 μm resolution. The Cy3 dye was excited at 532 nm and the Cy5 at 633 or 635 nm depending on the laser available. The fluorescent light was collected at 550-610 nm (Cy-3) and 650-750 nm (Cy-5). Self spotted and Agilent microarrays were scanned using Microarray Scanner G2565AA enabled by the Scan Control software (Agilent). Two scans were performed at 10% and 100% sensitivity of the photomultipliers. The two image files were processed by the Agilent Feature Extraction Software using an eXtended Dynamic Range (XDR) function to generate data with high resolution (potentially 5 orders of magnitude).

The NimbleGen arrays were scanned using the Agilent scanner for the control hybridizations and the GenePix 4200A microarray scanner (Axon Instruments) for the biological experiments. The scanning process was controlled by the GenePix Pro software and the data were extracted using the NimbleScan software (Roche NimbleGen).

Microarray Data Analysis

Raw data analysis was performed by Sven Wichert (MPI for Experimental Medicine, Göttingen) with the help of “R” statistical computing environment (<http://www.r-project.org>). The overall intensities of the microarrays were normalized using the RMA algorithm implemented in the R package. The average probe intensities were calculated across 4-8 replicate features depending on the design of the array. Average Cy5 signals represented the relative EXT expression levels and were normalized to the Cy3 signals originating from the plasmid DNA input. Plasmid input/Cy3 normalized EXT/Cy5 expression data were either given as relative differences (fold-changes) between samples or as arbitrary normalized expression values.

Synthesis and cloning of the EXT reporter library

EXT library synthesis

The synthesis of the EXT oligonucleotide library was carried out by the DNA Core Facility (MPI for Experimental Medicine, Göttingen) using standard DNA oligonucleotide synthesis chemistry. To create a complex oligonucleotide mixture, the synthesis followed a mix and divide strategy. After the synthesis of the 15-mer 3' invariable region, the sample was split into 8 portions each continuing the synthesis with a different sequence of 4 nucleotides. We have defined 8 specific 4-nucleotide combinations so called “words” that were used in this synthesis: CTTT, CAAA, ACAT, TCTA, TACT, ATCA, TTAC, and AATC. The resins carrying the oligonucleotides were mixed and subdivided again to continue with the 8 words at the next position. The cycle was repeated to introduce 5 randomly incorporated words in a row. Then a symmetrical 9-mer core sequence was generated by splitting the samples into 2 equal portion after each next nucleotide (WSWSSWSW, where W=A/T and S=C/G). The synthesis was completed by 5 additional words and a 5' invariable region. The resulting mixture of the 77-mer oligonucleotides was purified by HPLC and provided at a concentration of 50 pmol/ μl . The synthesis strategy is illustrated in **Figure 6** in the Results section.

Cloning of the pENTR EXT library

The EXT oligonucleotide library was diluted 10^4 times and PCR amplified for 30 cycles with the primers #10244 and #9754 complementary to the 5' and 3' invariable regions. The PCR product was then reamplified for 10 cycles with another primer pair (#10242 and #9789) to introduce the B3 and B2 recombination sites. Resulting PCR products were used for the generation of the pENTR shuttle clone library by recombination cloning. The BP reaction was set up in 20 μ l volume containing 4 μ l pDONR P3 P2 vector, 12 μ l PCR product and 4 μ l BP clonaseII enzyme mix. After an overnight incubation at 25°C the recombination product was purified by phenol chloroform extraction and ethanol precipitation, resuspended in 5 μ l H₂O and electroporated into 100 μ l undiluted DH-10b bacterial. The bacteria were allowed to recover in 2ml SOC medium for 1 hour and 1/20th of the bacterial suspension was plated to estimate the clone numbers. Remaining bacteria were transferred into 200ml LB medium, 50 μ g/ml Kanamycin and allowed to grow at 37°C for 4 hours. The culture was then centrifuged and the pellet was used for a midi-prep plasmid DNA preparation (Qiagen). The clone number in the preparation was estimated to roughly 100.000 clones. Several hundred colonies were screening by colony PCR with pENTR_s (#10977) and Dec2 (#9752) primers. The PCR products 280bp in size were separated on 3,5% agarose gels and 50% of the clones were discarded due to reduced insert length. From the other 50% of the clones some of the unpurified PCR product was diluted 1:10 with water and sequenced using pENTR_s primer (#10977).

After sequence verification 50-60% of the clones were selected that carried perfect full length inserts. For those clones mini-prep plasmid DNA was prepared and the bacterial stocks were generated in 96-well format. The overall cloning success rate of the entry clone EXT library should be considered around 25%.

Cloning of the G5 EXT library

The entry clone EXT library (L3/L2) in the form of the midi-prep DNA was subsequently used in an LR reaction with the G5-element (L1/L4), the TK promoter (R4/R3) and the pDEST_GL3_Luci to generate a library of GV responsive EXT reporter constructs (G5_TK_EXT-lib). The recombination reaction was set up in 20 μ l with 40fmol of the destination vector and 20fmol of each of the entry clone constructs and incubated at RT overnight. The LR reaction was purified by PEG precipitation and transformed into 100 μ l DH-10b. All the bacteria were transferred into 200ml LB/ampicillin and grown for 4 hours at 37°C. Different amounts of the suspension were plated out to estimate the density and the rest of the suspension was kept at 4°C overnight. The next day the bacteria were plated on square (25x25cm) agar plates containing 50 μ g/ml Carbenicillin and cultured overnight. The colonies were picked and transferred into 96-well plates containing 100 μ l per well LB/carbenicillin by the QPix Colony Picking Robot (Genetix). The bacteria were grown overnight and frozen at -80°C. The clones were screened by colony PCR using 0,5 μ l of the culture as a template. The inserts were amplified with the Dec1 (#10243) and Luc_as (#10764) primers and the PCR products were sequenced with the antisense primer (Luc_as, #10764).

About 2000 colonies were screened and 60% were discarded after the PCR analysis due to the size problem (30%) or failed PCR (30%). Among the remaining 40% only 18% were found to be perfect. 33% had non unique sequences, 27% failed in the sequencing reaction and 23% were missing 1-6 nucleotides. Final cloning success rate of the G5_TK_EXT-library can thus be considered to be at around 6%.

Cloning of the Cis-element EXT reporter constructs

Cis-regulatory elements were amplified from the available Luciferase reporter constructs along with the minimal promoter and a part of the 5' plasmid sequence to generate the *Cis*-element EXT reporter constructs. Depending on the template, different primer pairs were used carrying B4r and B3r recombination sites. Most of the *cis*-elements were derived from the Clontech's Pathway Profiling System (primers: #11059, #11055) and Panomics reporter constructs (primers: #11060, #11056). Each of the *cis*-element entry clone was recombined with three different sequence verified pENTR_EXT plasmids in one reaction. A total of 6 fmol of the EXT entry clone was used in recombination which makes 2 fmol per each each EXT. For each *cis*-element, 6-8 clones were sequenced to isolate the clones with different EXT sequences.

Transfections for Microarray Experiments

Transfections for microarray experiments were performed with adherent cells or trypsinized cells in suspension. In the latter case the cells coming from several different transfections were mixed and cultured under identical experimental conditions. For every microarray 10^6 cells were transfected with 5ng of each of the EXT reporter constructs. From 20 to 100 different plasmids were used in one transfection. For GV titration experiment as little as 2ng DNA per construct was used for the EXT pull exposed to the constant amount of GV. Reporter plasmids at mini-prep DNA quality were premixed and used for transfection to minimize the pipetting errors. Plasmids required for the protein-protein interaction assays were derived from midi-prep DNA preparations and 100 ng of DNA was used per 100 000 cells for each array.

List of cis-elements and associated EXT reporters

Cis-element	Min Prom	EXT Number	EXT sequence
NFkB	TATA	u109	AATCAATCCAAACTTTCTTTTGACGGTCACTTTTACATTCTACTTTTCTA
NFkB	TATA/TAL	u110	ATCAACATCTACTTTTCTAAGAGCGAGATCTATACTTCTATTACACAT
NFkB	TATA/TAL	u111	TACTACATACATACATTCTATCAGCTCTTCTAATCACTTTCTTTTAA
CRE	TAL	u112	ACATCAAATTACCAAATCTATCAGGGACATCTACAAAACATCTTTTAC
CRE	TAL	u113	ACATTACTTCTAAATCACATACTGGGTCTACATCTTTCTTTCAAATTAC
CRE	TAL	u114	AATCTACTTACTATCACTTTAGAGGCTGTCTTTTCTACTTTACATTTAC
E-box	TATA	u115	CAAAATCAACATATCATCTAAGAGGCTGTCTAAATCATCACAAAATCA
E-box	TATA	u116	TACTAATCTCTAAATCCTTTTGTCGGTGACTTTTACTCAAACCTTAAATC
E-box	TATA	u117	TCTACTTTACATCTTTACATACAGCCTCTACATTCTAATCATTACTACT
-	TATA	u118	ATCAATCATTACAATCCTTTTGAGCGACACTTTCTTAAATCAATCTACT
-	TAL	u119	CAAAAATCCAAATCTAACATACTCCCTGTACATTCTAATCATCTAACAT
IRF-1	TATA	u121	AATCTACTCTTTACATCTTTTGACCGAGTCTTTCTTTTCTATTACTCTA
IRF-1	TATA	u122	TACTACATCTTTATCAACATACTGGCTGTACATTCTATACTTACTTTAC
Stat3	TATA	u123	CTTTACATACATTCTATCTATGACGGAGTTCTAATCATACTTTACCAA
Stat3	TATA	u124	CAAAACAAATCTATCTAACATACACGGTCTACATTACTTACTTTACAATC
Stat3	TATA	u125	CAAATTACACATCTTTTCTATGAGCCTGTCTACAAAATCAATCATCTA
GAS	TATA	u126	TTACACATCTTTCTTTTACTCCGTGTCTTTCTTTTATCATCTA
GAS	TATA	u127	CAAACCTTTATCAATCACAAAACCTCGGTGACAAAACCTTTATCAACATCAA
GAS	TATA	u128	ACATAATCTCTAACATTCTAACAGGGTGTCTACAAAATACTTTACTACT
ISRE	TATA	u129	ATCATCTACAACTTTACATAGACCGACAACATTCTATACTTCTATACT
ISRE	TATA	u130	AATCATCATACTTCTAACATTGAGGCAGAACATTACTTCTATACTCTTT
ISRE	TATA	u131	TTACCTTTACATAATCACATAGTCCCTCAACATTTACAATCACATTTCTA
NFAT	TATA	u132	ACATATCATACTTCTACAAATGACCGTGACAACTTTACATTACTTCTA
NFAT	TATA	u133	ATCAACATCTTTACATCTTTAGAGGGACTCTTTTACTTACTCTTTACAT
NFAT	TATA	u134	AATCAATCTCTATCTATCTATGTCCACATCTATACTTCTATTACCAA
Myc	TATA	u135	ATCAATCATCTAATCACAAAACCTGGCTGTCAAACATACATAATCTTAC
Myc	TATA	u136	ACATACATTTACCAAACAAATGTGCGTGTCAAACAAATCTACTTTTAC
Myc	TATA	u137	CTTTTACCAAACAAATCTAAGAGGCTGATCTATCTACAAAATCTTAC
p53	TATA	u138	TTACTACTCTTTTACTCTTTTCACCGTCACTTTTACTTCTATCTACAAA
p53	TATA	u139	AATCCTTTACATCAAACAAATCAGGGTGTCAAATTACACATTACTTACT
p53	TATA	u140	TACTTTACACATATCACTTTACAGCGAGACTTTTACTTACTACTATCA
Rb	TATA	u141	TTACAATCATCACTTTCAAACCTCCGTGACAAAATACTATCATACTATCA
Rb	TATA	u142	ATCAAATCTACTACATCAAACCTGGGTCAAACTTTTCTATTACCAA
Rb	TATA	u143	ATCATCTAATCAAATCACATTGAGGCATACATAATCCTTTCAAATAATC
AP1	TAL	u144	TCTAACATCTTTACATCTTTACTCGGACACTTTCTTTTACATCATACT
AP1	TAL	u145	CTTTATCAAATCCTTTTCTATCAGCTCTTCTAATCACTTTACATATCA
AP1	TAL	u146	AATCTACTTACTATCACTTTTCAGCCAGTCTTTACATTACTTCTACAAA
GATA	TATA	u147	TCTAACATCAAACAAACTTTTCACCGAGACTTTAATCACATTACTCTTT
GATA	TATA	u148	CTTTATCAACATAATCCAAATGTGCGACTCAAATCACAAAACAAACAT
GATA	TATA	u149	ATCATTACTCTATACTACATTGAGGCTGTACATTACTCAAACAACTTT
HRE	TATA	u150	TCTAATCAACATAATCACATTGAGGGACTACATTACCTTTTACTCTA
HRE	TATA	u151	ACATATCATTACATCAACATTCTGCCTCAACATTACTATCAACATACAT
HRE	TATA	u152	TACTAATCTCTAAATCTCTAAGTGGGTGATCTACAAAATCAACATTCTA
E2F	TATA	u153	CTTTTCTATCTAAATCACATACAGCCTCTACATTCTAAATCATCAAATC
E2F	TATA	u154	TCTACAAATCTATCTACTTTTGACCCAGTCTTTTCTAAATCACATAATC
E2F	TATA	u155	TACTCTTTACATTACTACATACTCCGTCAACATACATTCTACTTTACAT
GRE	TAL	u157	ACATACATCTTTCAAACCTTTACACGCTGACTTTCAAATCTATTACTCTA

GRE	TAL	u158	TTACTACTTTACCTTTTCTATCACCGAGATCTACAAAACATTACTACAT
HSE	TAL	u159	CAAAATCAATCAACATACATTGAGGGTCTACATTACTTTACACATACAT
HSE	TAL	u160	TTACCAAATACTTTACACATACACGGAGTACATATCATCTATACTTACT
SRE	TAL	u163	ATCATCTATACTACATACATACACGCTCAACATTACTATCAATCAACAT
SRE	TAL	u164	TACTATCACAACTTTACATAGAGCCTCAACATCTTTACATTACTCAAA
MEF-1	TATA	u165	AATCACATAATCTCTAACATTGTGGCACTACATTCTAATCAACATTTAC
MEF-1	TATA	u166	ATCATTACAATCACATACATTCTGGCTGTACATCAAATCTATCTATACT
MEF-1	TATA	u167	TACTAATCCTTTTCTATCTATGACCCTGTCTATTACAATCCTTTATCA
MEF-3	TATA	u171	TTACTCTACTTTAATCACATTGTGGCTCTACATACATTCTAACATACAT
MEF-3	TATA	u172	TTACTACTTCTATTACTCTATGAGGGTCTAATCAATTTACTCTAACAT
MEF-3	TATA	u180	ATCATCTACAAAACATCAAATCTGGGACACAAAATCCTTTCAAAAACAT
ELK1	TATA	u174	TTACATCATTACTTACACATAGAGGGTCAACATTTACCTTTATCATACT
ELK1	TATA	u175	ATCACTTTTCTAAATCACATTCTGGGTCTACATTCTAACATTACTATCA
RxR	TATA	u177	CAAAATCATCTAATCATCTATGTGGCAGATCTAATCATACTACATCTTT
RxR	TATA	u178	AATCACATTTACACATTCTAAGTGCCAGATCTACAAAATCAAATCTTAC
RAR	TATA	u181	ACATTACTTTACATCATCTATGTGGTGATCTAACATATCAATCAAATC
RAR	TATA	u184	CTTTACATAATCCTTTCAAATCTGCCTCTCAAATACTTCTATCTACTTT
RAR	TATA	u185	ATCACAAATCTATACTTCTAACTGGGTCTTCTATCTAATCAAATCACAT
EGR3	TATA	u186	AATCTCTACAAAATCAAATGAGGCACCTCAAAAACATCAAATTACATCA
EGR3	TATA	u187	ACATATCACTTTAATCACATTCTAGGGACTACATCAAATACTCTTTAATC
EGR3	TATA	u188	ACATACATTACTTCTAACATAGTCCCTCAACATTTACCAAATCTAATCA
PR	TATA	u189	ACATTTACCTTTTTTACACATTGACCGTCTACATTACTTTACCTTTAATC
PR	TATA	u190	CTTTTACTACATAATCACATTGAGCCACTACATTCTATCTAACATCAAA
PR	TATA	u191	CAAATTACTCTACAAAATCTAACACCCTGTTCTATCTATCTAACATATCA
ER	TATA	u193	CTTTACATTCTATTACTCTAACTCCCAGTCTATACTTTACAATCACAT
AR	TATA	u195	CAAATACTAATCCTTTTTCTATCAGGGAGATCTATACTACATTACTTACT
AR	TATA	u196	ACATTACTACATTACTACATTCTCCGAGAACATTACTAATCTTACCAA
AR	TATA	u197	AATCATCATCTATTACCAAATGTGCGACTCAAAAATCAAATCTTACCAA
RxR2	TATA	u198	ACATTCTAACATTACTCTTTAGTGGTCACTTTACATATCAATCATTAC
RxR2	TATA	u199	CAAAACATCTTTAATCACATTGTGGCTCTACATCTTTACATAATCACAT
RxR2	TATA	u200	AATCATCAAATCACATACATTGTGCGAGTACATTCTATTACCTTTTTAC

example G5_EXTs

G5	TATA	u213	ACATACATTACTTACACATTCTGGGTGAACATTTACAATCAATCCTTT
G5	TATA	u214	CAAATACTACATCAAATCTATCACCGAGATCTATACTAATCATCATCTA
G5	TATA	u215	TCTAACATTACTATCACTTTAGACCCAGACTTTTTACTTACCAAATCTA
G5	TK	u102	TACTAATCCTTTCAAACCTTACACGGACACTTTACATCAAATACTAATC
G5	TK	u104	ACATTCTACAAAATCAACATAGAGGCTGTACATACATTACTATCAACAT
G5	TK	u62	CAAATCTATACTCAAAACATAGTGGGTGAACATTACTTACTATCATCTA
G5	TATA	u129	TACTACATTTACAATCTCTAAGTCCCAGATCTATCTACTTTTACTACAT
G5	TATA	u220	ATCATTACTCTATCTACTTTACTGCGTGACTTTCTTTTTACCTTTACAT
G5	TK	u93	TTACCTTTACATAATCACATTGTGCCTGAACATCTTTTTACTCTATACT
G5	TK	u94	ACATACATTCTACAAAACATACTCGCTCTACATAATCATCAACATTACT

RESULTS

The principle of EXTassays

EXTassays are based on the combined use of various types of reporter gene assay formats to quantitatively monitor very different molecular events simultaneously using one decoding microarray (**Figure 5**). To achieve multiplexing, the standard reporter gene is replaced by small unique oligonucleotide reporters, termed *expressed unique sequence tags* (EXTs), whereby each unique EXT serves as a specific reporter for a defined cellular assay. Multiple EXT reporters expressed in one cell or cell population can be isolated as a pool, amplified and analyzed simultaneously by hybridization to a diagnostic microarray (“barcode array”). Measuring the expression levels of all EXT reporters in this way can principally provide quantitative data for a plethora of cellular assays at once. The approach can easily be applied particularly for a large scale measurement of *cis*-regulatory element activities as each *cis*-element is physically linked to a particular EXT sequence in one plasmid construct. A mixture of such constructs can be transfected and assayed simultaneously since there will be no crosstalk between the “assays” operating in one cell. Activation of certain *cis*-regulatory elements will not only reveal the activity of transcription factors or transcription factor complexes but also indicate the activation of certain signal transduction pathways ([Romanov, Medvedev et al. 2008](#)). One of the advantages of this approach is that it provides a platform for the integration of other reporter gene coupled assays, i.e. for the measurement of signaling events upstream of transcription. For example, Split-TEV assays monitoring regulated protein-protein interactions can be coupled to a reporter gene readout and can thus be integrated into multiplexed EXTassays to measure signal transduction at different levels within one experiment ([Wehr, Laage et al. 2006](#); [Wehr, Reinecke et al. 2008](#)).

The ultimate goal is to provide a more complete view to the dynamics of the signaling cascade of interest. In the present setup, unambiguous coupling of each Split-TEV assay to a unique EXT reporter is ensured by coexpression of the assay components and an appropriate reporter within one cell. Therefore, separate transfections are performed for each protein-protein interaction assay and the cells are subsequently mixed and cultured under identical experimental conditions (**Figure 5**).

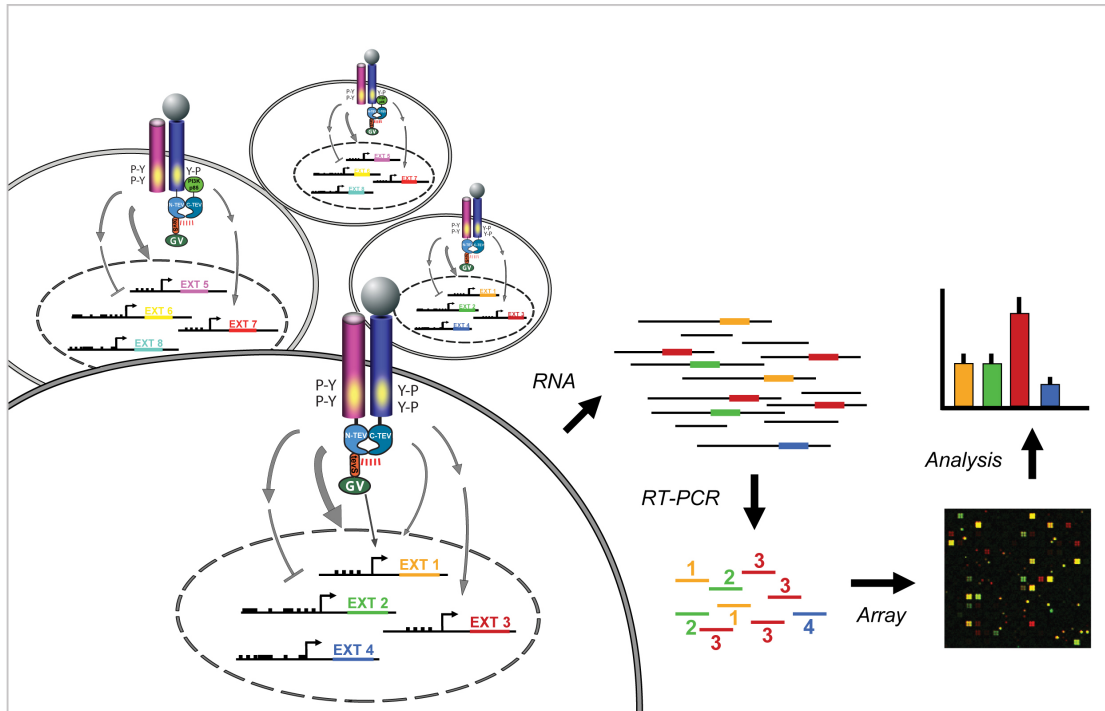


Figure 5. The principle of the EXTassay. EXTassays are based on simultaneous expression of multiple reporter constructs. These are invariably linking defined cellular assays to unique oligonucleotide sequences that can be discriminated e.g. by hybridization to a custom “decoding” microarray or by any other RNA profiling technique. Quantitative and simultaneous measurement of very different signaling events can be performed with different types of reporter gene assays. For example, Split-TEV assays can be combined with *cis*-regulatory assays to simultaneously monitor for receptor activation and downstream signaling at the level of activated transcription factors.

For EXTassays, total RNA is extracted and cDNA is synthesized. Subsequently, EXTs are amplified by PCR, labeled and analyzed with microarrays. The relative signal intensities of the different EXTs serve as a measure of the corresponding cellular assays. The performance of our approach critically depends on the functionality of the EXT oligonucleotide reporters. Unoptimized “random” sequences would lead to a distortion of quantitative data due to inhomogeneous performance of the reporters in amplification and hybridization procedures. All EXT oligonucleotides should display equal melting temperatures and no high energy secondary structure to allow optimal microarray hybridization performance. The library should contain a large number of unique EXTs to allow future upscaling of the EXTassay for high throughput applications.

Therefore, we have designed and synthesized a highly complex oligonucleotide library with the described properties. The combinatorial synthesis followed a mix-and-divide strategy illustrated by **Figure 6** ([Rossner 2008](#)).

Structure and properties of the EXT reporter library

The EXT oligonucleotide library contains 5' and 3' invariable regions and a central 49-mer variable region. The variable region has a symmetric organization and consists of 10 structural modules called “words” (W) that flank a central core region. Each word consists of 4 nucleotides. Out of all possible 4-nucleotide combinations only 8 are used that are built of three adenosine (A) or thymidine (T) residues, only one cytosine (C) and no guanine (G) residues at all. This structure yields balanced melting temperature and virtually absent intra- and intermolecular complementary regions for each member ([Brenner, Williams et al. 2000](#)). Each of the eight possible words (CTTT, CAAA, ACAT, TCTA, TACT, ATCA, TTAC, and AATC) occurs randomly at each of the ten possible positions. According to this principle, 8^{10} different combinations can be produced. The core region comprises nine bases of alternating A,T (W) or G,C (S) residues with three central G,C (S) residues (**Figure 6**) ([Rossner 2008](#)). Therefore 2^9 (2 bases at 9 positions) = 512 different core sequences are possible. The total complexity of the EXT library can be calculated as:

$$\begin{aligned}8^{10} \times 2^9 &= 1\,073\,741\,824 \times 512 = 549\,755\,813\,888 \\ &\approx 5,5 \times 10^{11} = \text{maximal library complexity}\end{aligned}$$

The 9-mer core element was introduced to allow monitoring the performance of each EXT with a few central mismatches and to enhance and stabilize the melting temperature profiles for hybridization approaches. Moreover, the core element increases the complexity compared to a pure word structure by approximately one order of magnitude. The balanced base pair composition and homogeneity of the EXT library was analyzed by Sven Wichert using R-scripts. A large virtual library of 100000 (100K) EXTs and an equal size library of random 49-mer sequences were generated and their melting temperature (T_m) profiles were calculated. These data were then compared to the T_m data for 229 cloned and sequence verified EXTs. The virtual and cloned EXT libraries display a highly narrow T_m range with an average of 61,2 °C and a standard deviation of 0,7 and 0,8, respectively. In contrast random 49-mers display a much broader T_m distribution and significantly higher average melting temperature of $70,1 \pm 3,3$ °C (**Figure 7a**). The EXT design also yielded in a low cross-similarity profile within the cloned EXT library. The majority of the 229 EXT oligonucleotides display only 20% similarity in pairwise comparison to each other (**Figure 7b**).

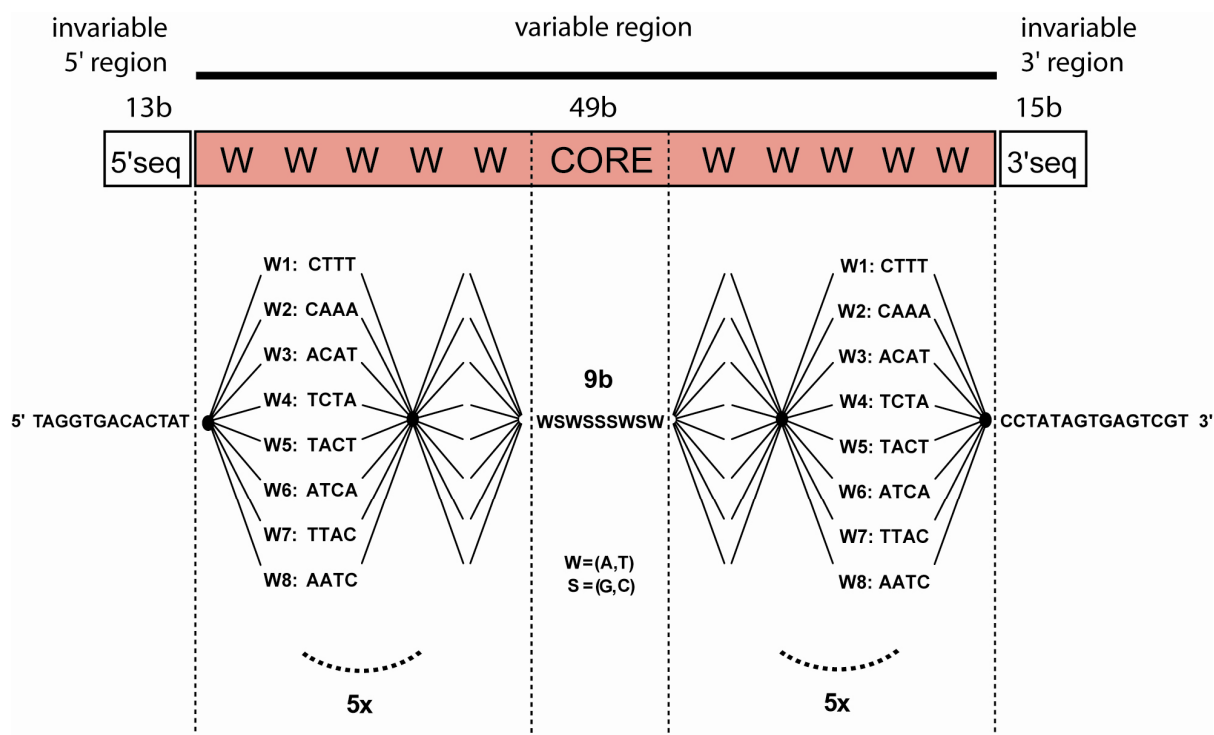


Figure 6. Structure and synthesis strategy of the EXT oligonucleotide library.

Each oligonucleotide in the library carries 5' and 3' short invariable regions and a 49-mer variable region consisting of ten 4-nucleotide units (“words”=W) and a 9-mer core sequence. The lower part of the figure illustrates the “mix and divide” principle of the EXT library synthesis used for random incorporation of the 8 possible words into 10 possible positions (a more detailed description is available in the Materials and Methods section).

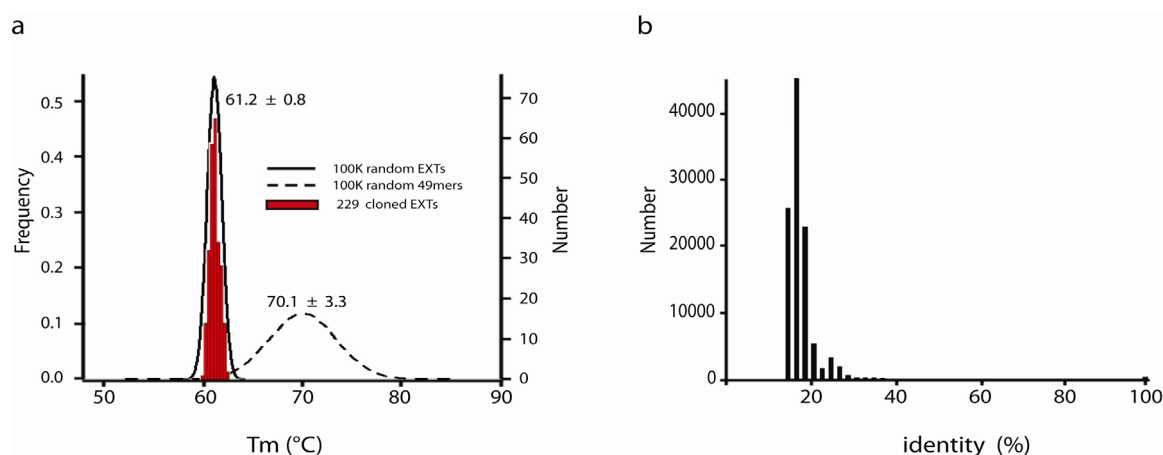


Figure 7. Melting temperature (T_m) and similarity profiles of EXT oligonucleotide libraries.

(a) The histogram in red shows the T_m distribution for 229 cloned and sequenced EXTs. It follows the narrow range outlined by the calculated T_m distribution for 10^5 randomly computed EXT sequences. All EXT oligonucleotides show a very consistent and narrow range of melting temperatures compared to 10^5 computed random 49-mer sequences (dashed line). For computed libraries the frequency values are shown which indicate the number of sequences with particular T_m relative to the total.

(b) Similarity profile histogram for a cloned library of 229 unique EXTs shows that the majority of sequences share less than 20% identity. The y-axis indicates the number of pairwise combinations with a given identity score.

Optimization of the hybridization conditions

All data on the mismatch control performance shown below are derived from hybridization of custom Agilent microarrays at 65°C with the supplied buffer supplemented with 15% formamide and blocking DNA for increased stringency. Hybridization of NimbleGen arrays were equally optimal at 50°C with standard hybridization components (data not shown).

The cross hybridization profile of the mismatch controls is a function of hybridization conditions. In general, an optimal trade off between high stringency and overall signal intensity should be achieved.

We have tested a range of hybridization conditions and hybridization temperatures. Three different microarray platforms were tested: self spotted microarrays and custom microarrays from Agilent and Roche/NimbleGen. NimbleGen arrays were finally used for the biological experiments. Hybridization of the NimbleGen arrays at 45°C resulted in substantially elevated levels of cross hybridization. At 60°C and 55°C the signal intensities were not sufficient to exploit the full range of the scanner dynamics. The best outcome was obtained at 50°C with high signal intensities for perfect match probes and adequate mismatch signal profiles. For the biological experiments later on, the temperature was reduced to 48°C to facilitate the resolution of lower signals. This has slightly increased the signals at the mismatch controls but did not affect the data quality since the extent of similarity between the unique EXTs used for the experiment is by far lower than for the mismatch controls and far beyond a critical cross hybridization score.

EXT performance in a microarray hybridization

Cross hybridization is one of the most important issues to be addressed when working with microarrays. The probes have to be selected for low similarity and the hybridization conditions have to be optimized to achieve high discrimination.

We have used a broad panel of control sequences to explore the EXT properties with respect to cross hybridization and to monitor the stringency under given hybridization conditions. We randomly selected five “control” EXTs (c1-c5) for a thorough evaluation of their performance in microarray experiments (**Figure 8**). For each of these EXTs, a set of controls was designed by introducing various numbers of mismatch nucleotides at different positions within the sequence. Roughly 700 different mismatch controls per EXT were used which makes 3500 in total. To introduce a mismatch without altering the melting temperature of the oligos, adenine (A) was always exchanged to thymidine (T) and cytosine (C) to guanine (G) and vice versa. The following mismatch controls were designed: (i) 1-8 bp long mismatches were introduced at every possible position along the “mother” sequences of c1-c5, (ii) different word mismatches or word mismatch combinations at different positions and (iii) combinations of single nucleotide mismatches in the core element with word mismatches.

Among the eight possible words used in the EXT library there are always 2 words that share a C at the same position and 3 words that share an A or T at the same position, however, not more than one nucleotide is identical. While G/C nucleotide pair contributes three hydrogen bonds to the hybridization the two words sharing a C have to be considered the most similar of all. To test whether such words would be discriminated, a subset of mismatch controls was designed where one or more words

EXT	W	W	W	W	W	core	W	W	W	W	W
c1	CTTT	ATCA	AATC	TACT	TCTA	TGAGCCAGT	TCTA	TCTA	TACT	TACT	ACAT
c2	AATC	TACT	AATC	TCTA	TCTA	TCAGGGACT	TCTA	AATC	AATC	TCTA	CTTT
c3	CAAA	ACAT	ATCA	TCTA	ACAT	TCTGGGTCA	ACAT	CTTT	ATCA	CTTT	AATC
c4	TCTA	TCTA	CTTT	TACT	CTTT	TCTGGCACT	CTTT	ACAT	TACT	AATC	ACAT
c5	CTTT	CTTT	TTAC	ATCA	ACAT	ACTCGCTCT	ACAT	TCTA	CAAA	CTTT	ACAT

Figure 8. Control EXTs c1-5. Five randomly picked oligonucleotides that were cloned from the EXT library and sequenced show low similarity to each other. Each of the 8 possible words is depicted in a different color to illustrate the diversity of word composition. These five sequences were designated control EXTs c1, c2, c3, c4, c5 and used later on as hybridization controls.

were exchanged to its most similar word-counterpart: CTTT/CAAA, ACAT/TCTA, ATCA/TACT, AATC/TTAC. Up to five word mismatches were introduced at various positions as well as combinations of single word mismatches distributed along the sequence.

Microarrays with mismatch control as well as perfect match oligonucleotides complementary to c1-c4 (the probes) were hybridized to the corresponding fluorescently labeled c1-4 products (target). Thus, we could precisely determine cross hybridization profiles and optimize the conditions to yield optimal hybridization dynamics. As expected, perfect matches displayed the highest fluorescence intensity. At optimized hybridization conditions mismatch sequences showed dramatically reduced signals depending on the position and length of the mismatch. The cross hybridization was estimated as a percentage of the fluorescence observed for a mismatch probe relative to the perfect match of the corresponding control EXT. As shown in **Figure 9**, a significant discrimination was observed already with one base mismatch at central positions. The discrimination was further improved as the length of the mismatch was increased to 2, 4, 6 or 8 nucleotides. This was, however, not the case for the sequences with the central region identical to that of the control EXT (**Figure 10**, **Figure 11**). The extent of discrimination depended to a larger extent on the position of the mismatch rather than on its length. Analysis of the word based mismatches showed that a word mismatch displays the same discrimination as a corresponding 4-nucleotide mismatch even though one of the four nucleotides of the word (always the C) was not altered (**Figure 11**). This means that all of the eight words can be discriminated equally well. One word mismatch displayed full discrimination when located rather centrally at position 3-7 but not at the outermost 4 slots. These positional effects could be abolished by having at least one base mismatch at any position in the core sequence (**Figure 12**). In other words, a perfect discrimination could be achieved with as little as one central base and one word difference at any position. Without a central mismatch, 5 word mismatches were needed to reach full discrimination independent of the mismatch position (**Figure 13**). For oligonucleotides with mismatches only at the boundary, the minimal difference in sequence required for full discrimination is approximately 40% (5 words = 20bp, $20/49=40,8\%$). For sequences with one or several mismatches at other positions far more similarity can be allowed. With the achieved hybridization specificity for the tested microarray platforms, the least possible complexity of the cross hybridization

free EXT library equals the complexity of the EXT central region without the 5 words at the boundary and can be calculated as:

$$8^5 \times 2^9 = 32.768 \times 512 = 16.777.216 \\ \approx 1.7 \times 10^7$$

= minimal complexity allowing full discrimination with microarrays

So far, we have cloned and sequenced about 600 unique EXTs and have not encountered any problem of EXT similarity. We conclude that the EXT reporter library is indeed physically highly complex and provides large numbers of significantly different sequences. As the size of the operating EXT library grows in the future, however, sequences with higher similarity will be emerging which should not be used in the same experiment. Without considering the positional effects, a general cut off could be set to a minimum of 40% overall sequence difference (see above) or maximum 60% identity. This would be a very rigorous requirement as the vast majority of excluded sequences would not show any interference. Theoretically, computer algorithms could be developed that would allow to also consider position effects to increase the number of EXTs available for experiments. This will, however only be needed when more than 10 million assays shall be performed in parallel. Therefore, when compiling new EXT sets by cloning and sequence verification the results of the control experiments described here have to be taken into account. It is important to note that the hybridization dynamics of all control EXTs is highly similar. The signals of the mismatch probes follow the same pattern of signal intensity which again points out the uniformity of the EXT library and its balanced structure.

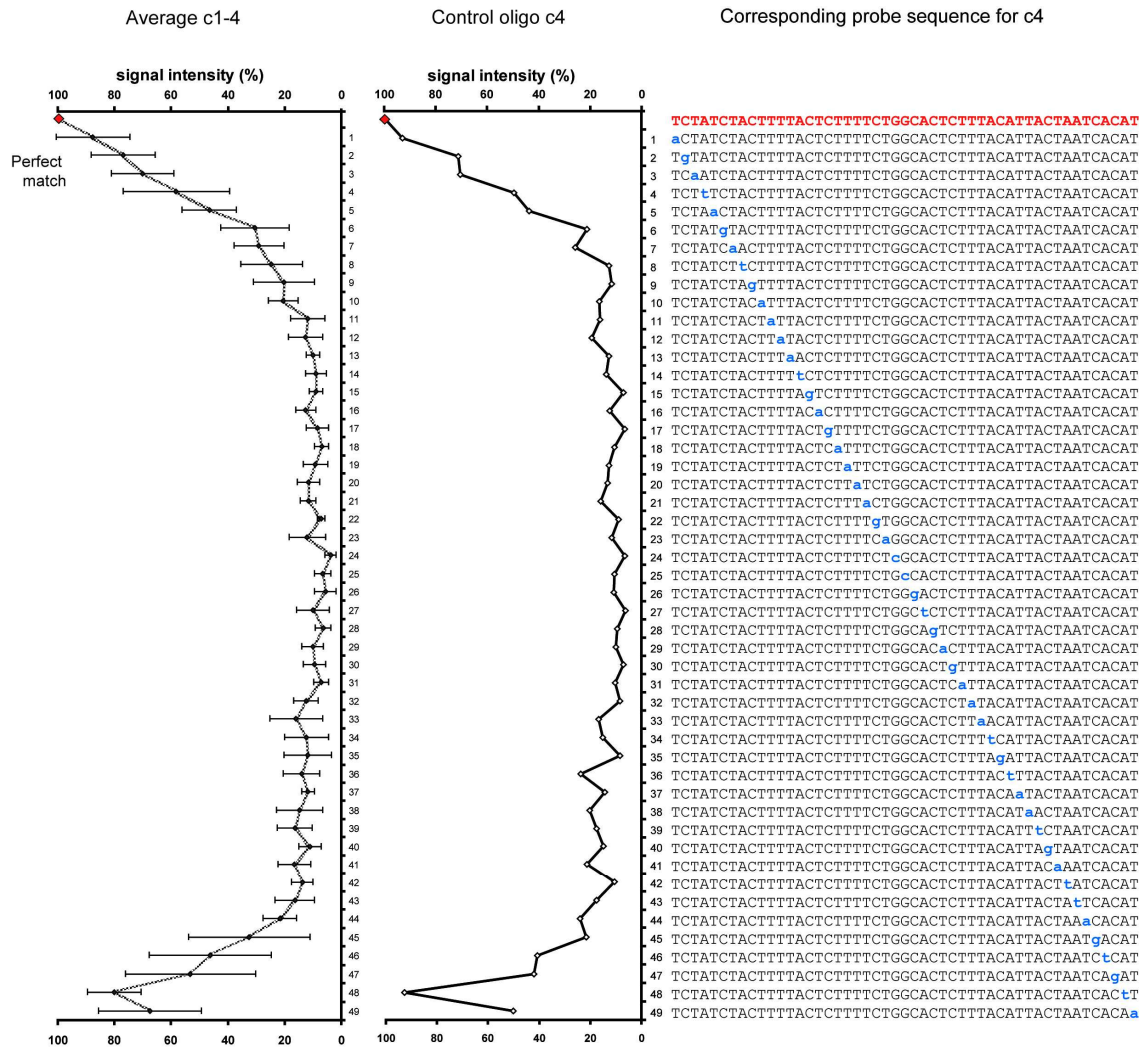


Figure 9. Cross hybridization profile for single base pair mismatch controls. Four control EXTs c1-4 were hybridized to a microarray harboring a perfect match and a set of mismatch sequences. Representative sequences for EXT c4 are shown on the right. The perfect match sequence is depicted in red and the position of the mismatch is highlighted in blue. Due to high sequence similarity, EXTs c1-4 cross hybridize to the mismatch sequences. The extent of cross hybridization is plotted for each mismatch probe as a percentage of the perfect match signal. Average cross hybridization profile for all 4 control EXTs is shown on the left and specific data for the c4 in the middle. Even with one base mismatch in the central region, the signal drops dramatically suggesting a high specificity of the hybridization. Nevertheless, with such minor sequence differences full discrimination can not be achieved.

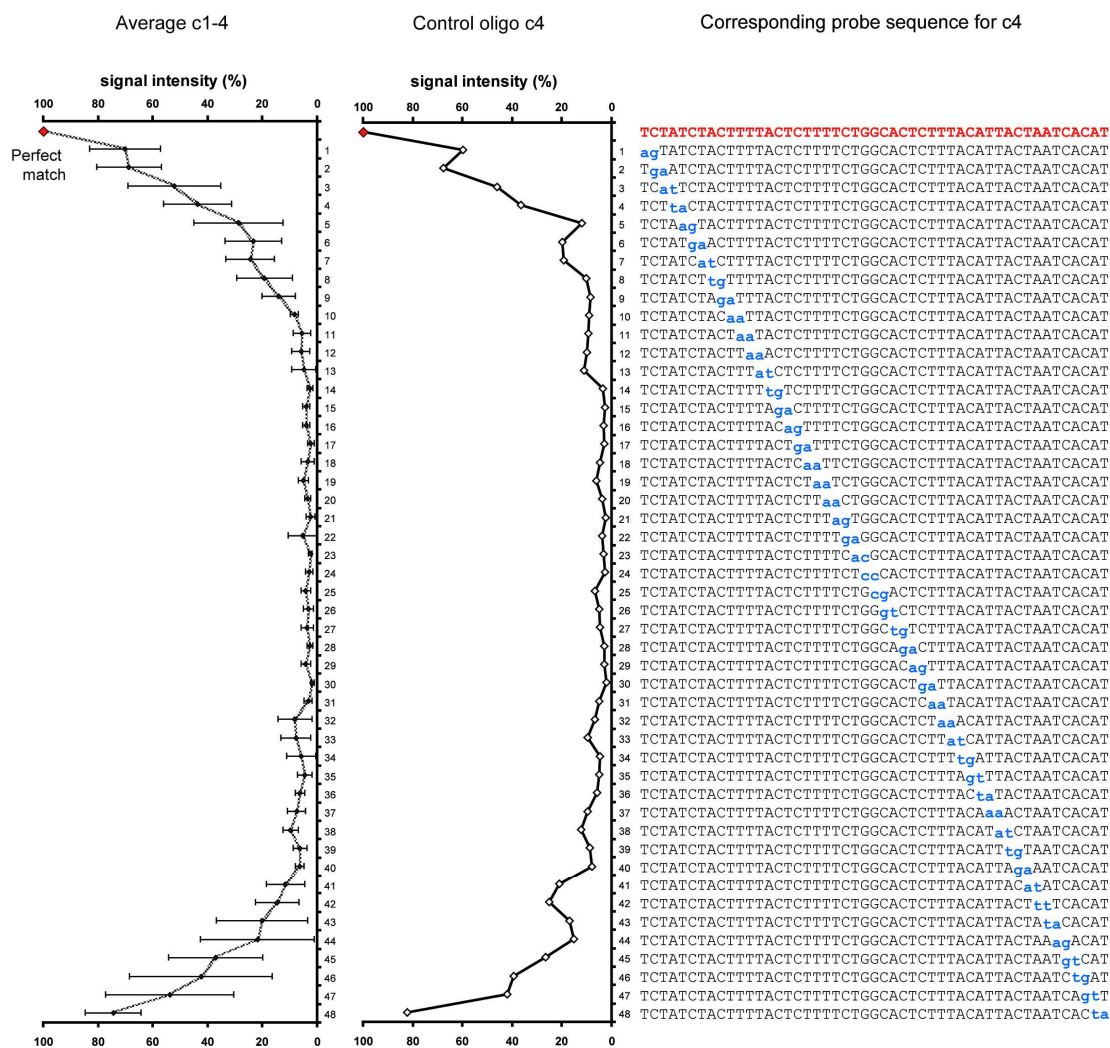


Figure 10. Cross hybridization profile for two base pair mismatch controls. Four control EXTs c1-4 were hybridized to a microarray harboring a perfect match and a set of mismatch sequences. Normalized spot intensities for EXT c4 are shown in the middle and corresponding probe sequences on the right. The average intensity plot of c1-c4 on the left shows highly reproducible behavior of all four control EXTs. All signals are normalized to the perfect match intensity. The discrimination is substantially improved compared to one base mismatch. The average signal in the central region (excluding 2 boundary words on each side) is 11 ± 4 for 1b and 5 ± 2 for 2b mismatch. For some central mismatches the signal drops to 2% of the maximal intensity close to the background level (1,65%) implying that even highly similar sequences can be fully discriminated at optimized hybridization conditions.

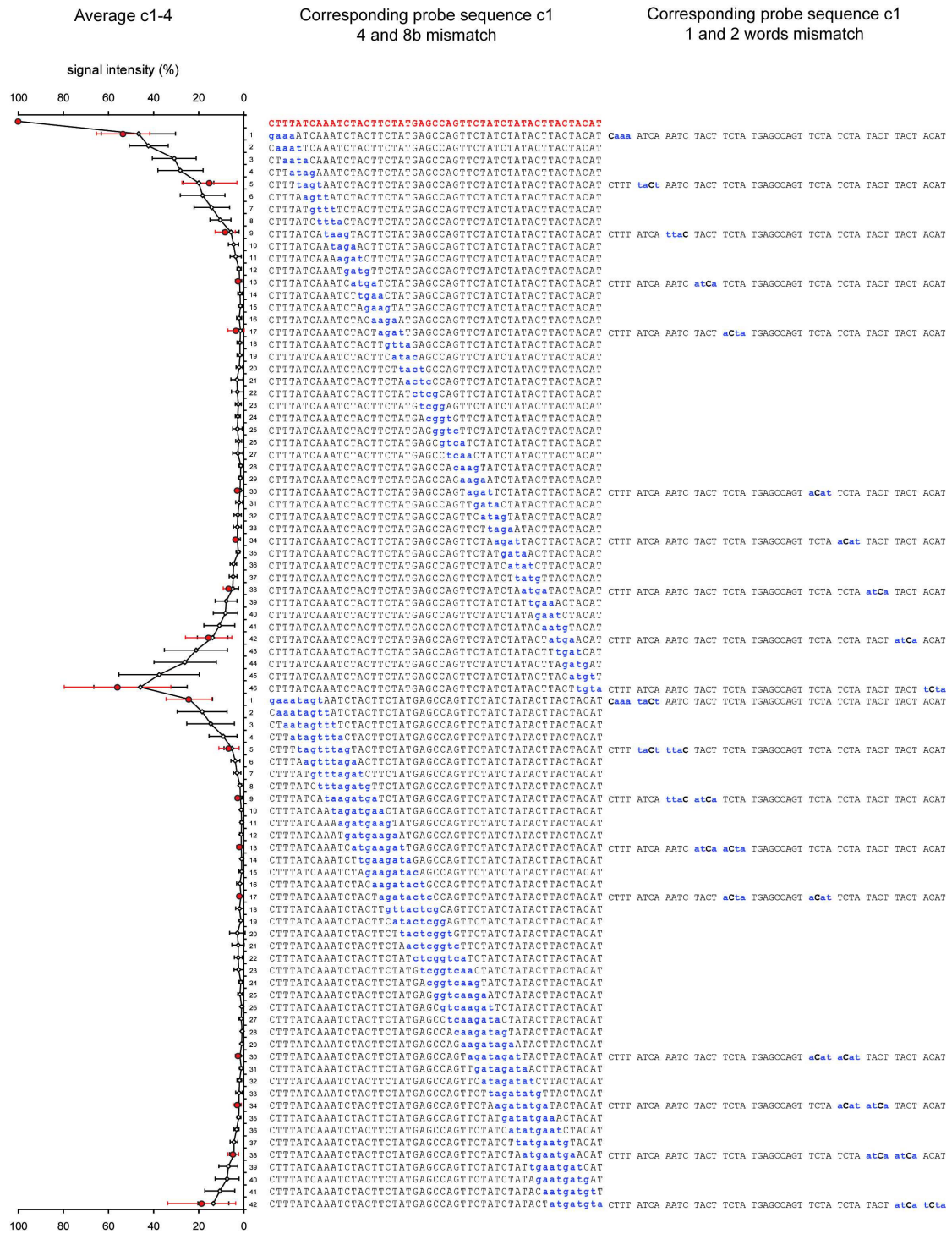


Figure 11. Comparison of the cross hybridization profiles for 4 and 8 bases versus 1 or 2 words mismatches at corresponding positions. Average data for c1-4 and sequences for c1 are displayed. Cross hybridization levels for 4 and 8 base mismatch probes are plotted in black and for 1 and 2 words mismatch probes in red. In contrast to a 4-base mismatch, in 1 word mismatches a cytosine base is not altered which should be considered because words with one common nucleotide are found in the EXT library. The signals appear to have very similar dynamics with only slightly higher edge effects for one word mismatch. Which means that all words can be well discriminated. For most of the mismatch sequences perfect discrimination is observed. Although some edge effects remain for EXTs with only one altered word, a reliable discrimination is achieved.

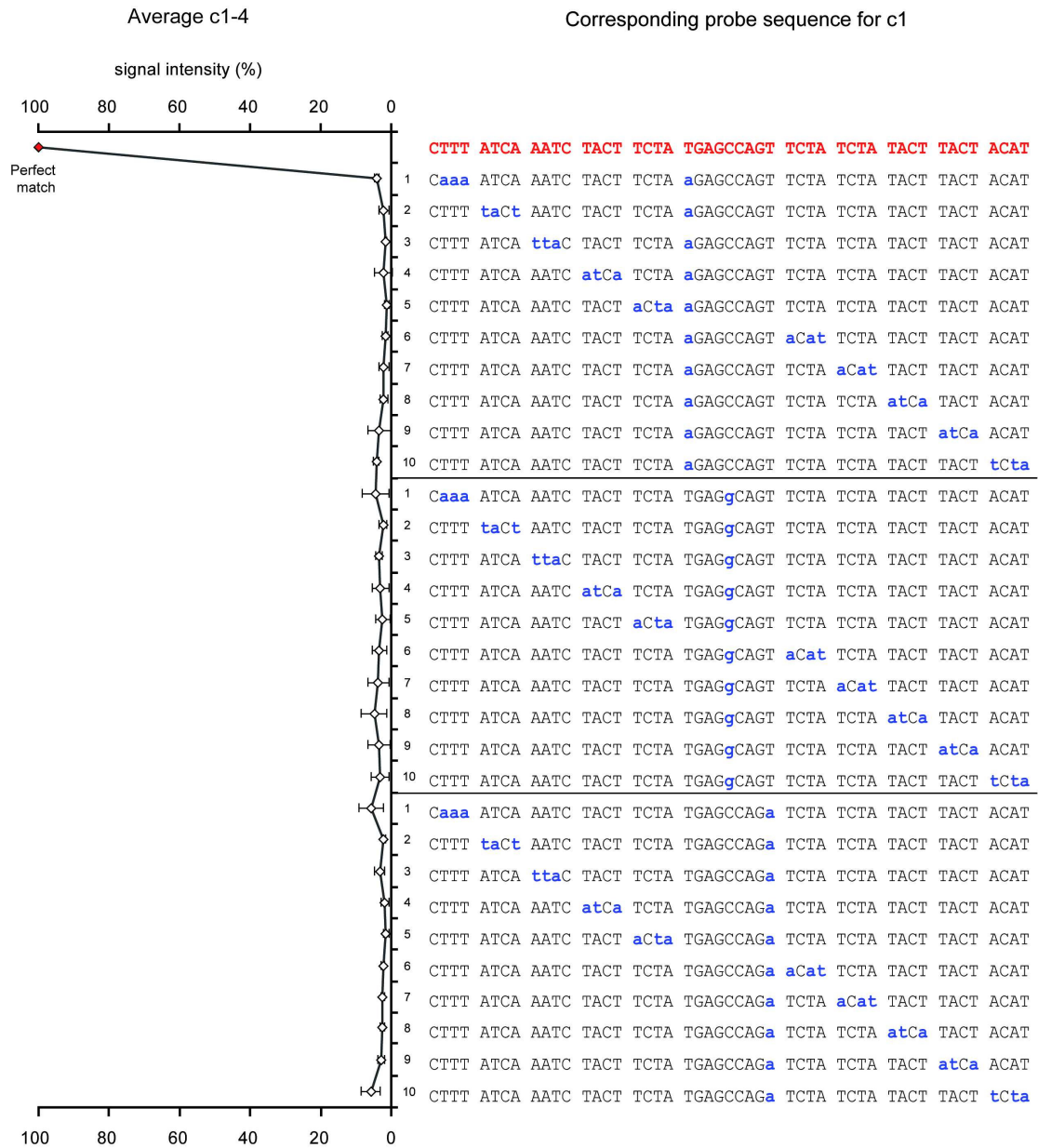


Figure 12. Combining single central mismatches with single word mismatches. Oligonucleotides carrying at least one base mismatch in the center and at least one word mismatch at any position can be fully discriminated by microarray hybridization. Boundary effects for one word mismatch controls are abolished by only one additional base mismatch in the center. The variation depending on the central mismatch or a word position is low.

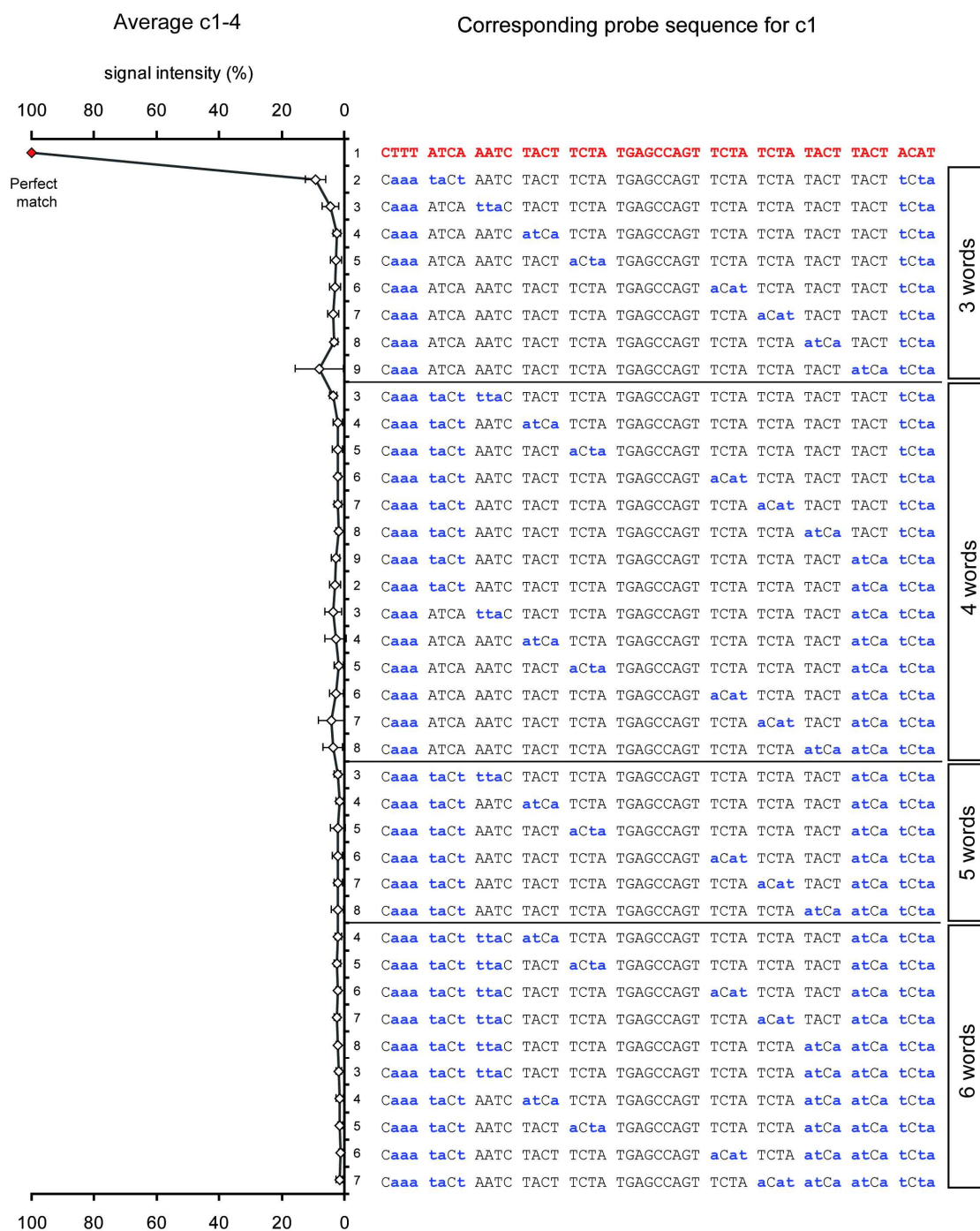


Figure 13. Cross-hybridization profiles with multiple word mismatches. To attain full hybridization specificity for the sequences with identical core region several words differences are required. With more than 4 words differences, perfect discrimination is achieved independent of the mismatch position. Sequences with boundary word mismatches are shown because at more central positions even higher discrimination is expected.

Dynamic range of the microarray based EXT measurement

In order to address the performance and dynamic range of the EXT microarray analyses, we have mimicked an EXT assay with a defined signal range by using a “spike in” mix of eight individually labelled unique EXTs. Two distinct EXTs were allotted to each of the concentrations ranging over two orders of magnitude. 31,2 fmol of the spike in mix (containing 0.1 fmol of the EXT13 and 14, 0,5 fmol - EXTs15/16, 5 fmol - EXTs17/18 and 10 fmol - EXTs19/20) were hybridized to a NimbleGen microarray in the presence of 85 fmol of the control EXT c5. In a second experiment, double the amount (2x) of the spike in mix and the same amount of the c5 was analyzed. Two fold differences of the spike in EXT signals over a range of concentrations were assessed (

Figure 14) The concentrations were adjusted to the amount of the Cy3 dye (fmol) and not to the amounts of RNA, to correct for slight differences in dye incorporation rates from EXT to EXT. In a real experiment all EXTs would be labelled in a complex mixture and therefore assumed to have equal labelling efficiencies.

All hybridized EXTs showed signals proportional to their concentrations in the mix. Already at 0,1 fmol, an increase in the signal intensity over the background was detected for EXTs 13 and 14 (178 ± 30). Two fold differences between the two hybridization mixes could be measured. The average fold difference between the spikes in the 1x and 2x mixes for all of the EXTs was only slightly below the expected two-fold difference ($1,84 \pm 0,19$). This slight deviation from the expected ratio is consistent for all eight EXTs and can, therefore, most likely be attributed to a pipetting error.

The average background value of roughly 100 (110 ± 5) was calculated from the spot intensities corresponding to the EXTs that were not present in the target mixture. The maximal signal of the particular hybridization is represented by the average c5 intensity at around 30000 (29117 ± 3722). The dynamic range outlined by these two values is about 300 and constitutes more than two orders of magnitude. As long as the c5 intensity does not reach the maximum of 65000 units allowed by the scanner sensitivity and the Cy3 dye is used that displays elevated background compared to the Cy5, this range could in fact be improved to roughly 600 for Cy3 dye and up to three orders of magnitude for Cy5. Two fold differences in the EXT abundances could be robustly measured at the amount of 0,5 and 1 fmol which relates to the overall amount

of the target as approximately 1:10 000 ($8500 / 0,5 = 17000$ and $8500 / 1 = 8500$). This indicates that with respect to target concentrations in a complex EXT mixture at least 4 orders of magnitude can be resolved.

In summary, the optimized microarray readout for the EXT quantification allows to monitor abundance differences within complex samples with high sensitivity over more than two orders of magnitude.

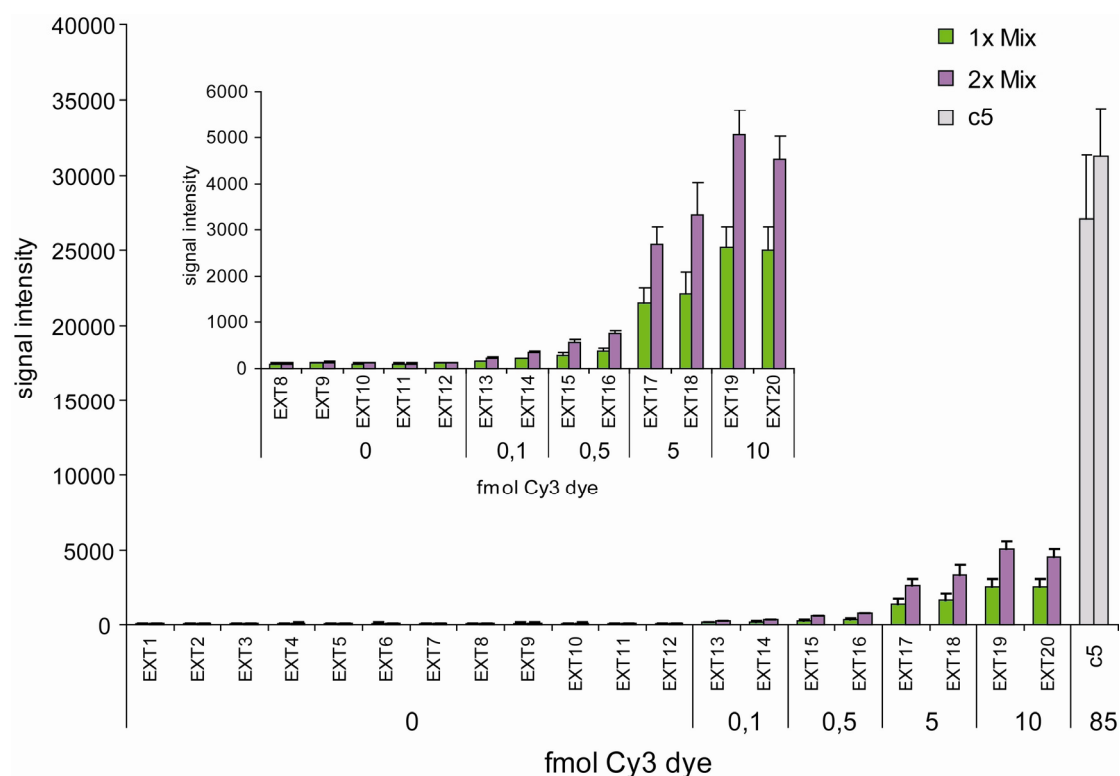


Figure 14. Assessment of the dynamic range of microarray based EXT analysis. A spike in mix of individually labelled unique EXTs 13-20 was prepared with individual EXT concentrations ranging over two orders of magnitude. 31,2 fmol of the spike in mixes (containing 2 different EXTs at the amount of 0,1fmol, 0,5fmol, 5fmol and 10fmol fluorescent dye) were hybridized to one microarray (1x) and double the amount (2x) to the second microarray. High excess (8500 fmol) of the control EXT c5 was added to both hybridizations. The signal profiles show increasing intensities for the EXTs spiked in at rising concentrations (inlay depicts EXTs 8-20 at reduced scale). The dynamic range spans over more than two-orders of magnitude. Average background represented by the signals of EXTs 1-12 is around 110 ± 5 and the average c5 signal reaches up to 29117 ± 3722 .

Robustness against amplification biases

Measurement of the relative abundance of the EXT reporters by microarray hybridization involves several amplification steps. Current protocol includes an exponential PCR amplification and two linear amplification steps: reverse transcription and in vitro transcription. It is known that due to different mechanisms depending on the template length, relative abundance and/or nucleotide sequence composition some templates tend to be unequally amplified in PCR reactions and most likely also in IVT reactions ([Kanagawa 2003](#)). As a result, the composition of a complex mixture may be biased during PCR amplification. The balanced structure and base pair composition of the EXTs should not only facilitate their optimal performance in the microarray hybridization but also make sure that the relative abundance of each EXT within the complex mixture stays constant during amplification procedures. To verify this assumption, we have performed a series of control experiment by comparing the relative abundance of amplified and unamplified EXT mixtures. Different amounts of control EXTs c1-4 were combined in the presence of a large excess of the c5. Four mixtures were prepared, each one containing twice more of c1-4 than the previous one (MixA:MixB:MixC:MixD=1:2:4:8). Therefore, each EXT should theoretically show a two fold difference from mix to mix when analysed by microarrays. The exact compositions of the mixes are shown in **Figure 15a**. All mixtures were diluted 10^4 , 10^8 and 10^{10} fold and PCR amplified for 40 cycles. According to the dilutions, 0,1ng, 10fg or 0,1fg were used as input. The PCR products as well as unamplified mixtures were labelled and analyzed by hybridization to self spotted microarrays (**Figure 15b,c**). Signals in the amplified and unamplified samples show highly similar patterns of intensity and the two-fold differences between the mixes A-D are preserved after the amplification. Average correlation coefficient between all four data sets is $0,991 \pm 0,014$. This strongly suggests that the PCR amplification does not introduce a detectable bias into composite EXT mixtures.

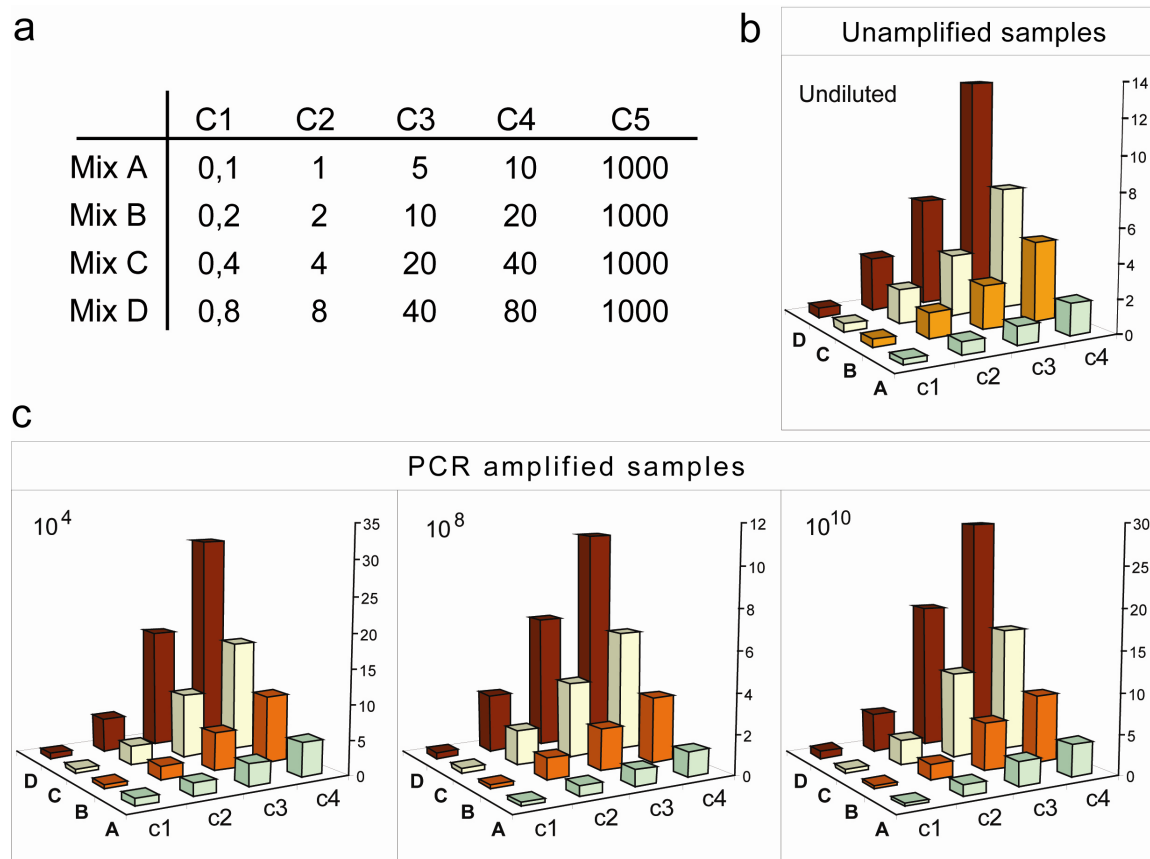


Figure 15. Controlling a potential PCR amplification bias. (a) Composition of the 4 control mixes (A, B, C and D). The amount of input PCR products for control EXTs c1-5 is given in nanograms. (b) Patterns of signals produced by the control mixes hybridized to 4 separate microarrays (A, B, C and D). The c5 signal is not shown for better visualization of lower signals. (c) Control mixes (A-D) were highly diluted, PCR amplified and analyzed by microarray hybridization. Dilution factors are shown in the corner of each graph. The pattern of signals is not altered after PCR amplification and is independent of the dilution factor which indicates homogenous and reproducible amplification.

In summary, the set of control experiments above has revealed that:

- (i) the hybridization dynamics of c1-4 are highly similar
- (ii) even very similar sequence-related EXTs can be discriminated with high specificity
- (iii) abundance differences can reliably be determined at very low input concentrations
- (iv) amplification of EXT template mixtures with PCR does not cause any detectable bias in the sample over ten orders of magnitude.

We conclude that EXTs are optimally suited targets for microarray analyses and should serve as valid reporters to reliably convey signals from cellular assay.

Design and analysis of EXT reporter constructs

The invariable regions of the EXT were introduced to serve as primer binding sites for amplification and subcloning purposes (**Figure 6**). The EXT library was amplified with primers carrying long overhang regions to attach sequences important for the functionality of the EXT reporter oligonucleotides in the experimental workflow (**Figure 16**). The outermost sequence stretches contained specific recombination sites for efficient cloning of resulting PCR products. Additionally, for the amplification of expressed EXT pools after reverse transcription, two primer binding sites designated as “Decoder primers” (Dec1 and Dec2) were introduced. And finally, bacteriophage T7 and T3 promoter sequences were placed adjacent to the EXT variable region to enable *in vitro* transcription from preferred strands to generate target RNAs for microarray hybridizations. To generate large numbers of EXT-reporter constructs, we used a multi-component modular recombination cloning strategy (“Multisite Gateway”) (**Figure 16**). Initially, a shuttle clone library was created carrying approximately 10^5 full length EXTs. About 500 single clones were isolated and sequenced. In parallel, thirty *cis*-regulatory elements (**Table 1**) and two minimal promoters: TATA-box and herpes simplex virus thymidine kinase promoter (TK) were cloned into similar shuttle clone vectors. Most of the *cis*-elements were amplified along with the minimal promoters “TATA”-box or TATA-like promoter (TAL). For *cis*-reporters, no additional minimal promoter was used and the constructs were 5’ supplemented with a strong transcriptional pause site (**Figure 17**). Final reporter constructs were generated by single recombination reactions of several selected shuttle clones into a firefly luciferase reporter vector backbone. In the final assembly, a *cis*-element, a minimal promoter and an EXT were placed upstream of the firefly Luciferase reporter gene such that upon *cis*-element activation, an EXT and the firefly luciferase are expressed on the same mRNA. Therefore the same constructs could be used to conduct multiplexed EXT assay and single luciferase reporter measurements to compare the two techniques and assess the reliability of the EXT assay approach. Three unique EXTs were assigned to each *cis*-element, to generate three distinct EXT reporter constructs of which at least two were successfully cloned and used in the experiments. The list of EXT reporters used in his study in association with the *cis*-regulatory elements is shown in the Materials and Methods section.

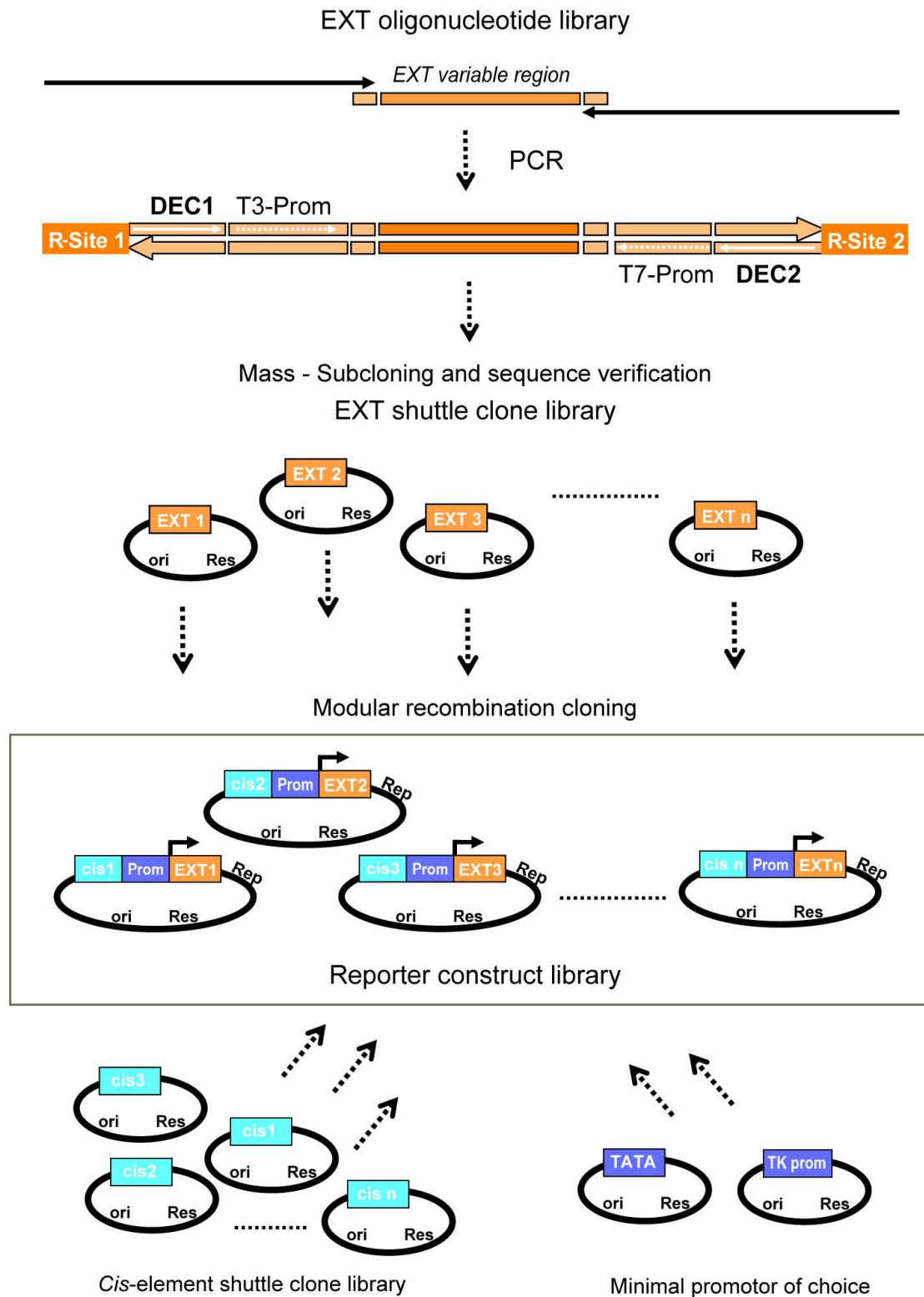


Figure 16. EXT amplification and subcloning strategy. The EXT oligonucleotide library was PCR amplified and subcloned to generate a library of shuttle clones. Several functional elements were introduced flanking the EXT sequence: recombination sites (R-Site1/2), “decoder” primer binding sites (DEC1/2) and viral promoters (T3/T7). Three shuttle clones harbouring different functional elements were used in one recombination reaction to place a *cis*-regulatory element, a minimal promoter and an EXT in front of the firefly luciferase reporter gene (Rep). Depending on the shuttle clone, defined *cis*-regulatory elements were linked to unique EXTs.

Table 1. List of *cis*-regulatory elements

<i>Cis</i>-element	Transcription factor	Pathway
MEF-1	MYOD-1	
MEF-3	Six1-6 homeoproteins	Muscle differentiation
ELK1	ELK-1	Ras/Raf/MAPK
EGR3	EGR3	MAPK
NFAT	NFAT	PKC & Ca ²⁺ /calcineurin
GATA	Globin TF family (GATA1-6)	BMP/ Calcineurin
HRE	HIF-1	Hypoxia
IRF-1	IRF-1	JAK/STAT
Stat3	STAT3/STAT3	JAK/STAT
GAS	STAT1/STAT1	JAK/STAT
ISRE	STAT1/STAT2	JAK/STAT
E2F	E2F family	cell cycle
Myc	Myc	Wnt, Shh, EGF, MAPK
p53	p53	cell cycle/apoptosis
Rb	Retinoblastoma protein	cell cycle
AP1	AP-1 (c-jun/c-fos)	SAPK/JNK /MAPK
CRE	CREB	JNK/p38 & PKA
GRE	Glucocorticoid Receptor	Glucocorticoides
HSE	HSF	heat shock response
SRE	Elk-1/SRF (TCF)	MAPK/JNK
NFκB	NFκB	NFκB/IKK
E-box	bHLH proteins	Metabolic and circadian control
PR	Progesteron Receptor	Progesteron
ER	Estrogen Receptor	Estrogen
AR	Androgen Receptor	Androgens
RxR	Retinoic X Receptor	Retinoic Acid
RxR2	Retinoic X Receptor	Retinoic Acid
RAR	Retinoic Acid Receptor	Retinoic Acid
G5	Gal4-VP16 (GV)	artificial transcription factor

A separate library of EXT reporters was cloned under the control of the “G5” *cis*-element (5 clustered Gal4 binding sites) that can be activated in mammalian cells by the synthetic transcriptional transcription factor GV (a fusion of the yeast Gal4 DNA binding moiety and the strong transactivator domain derived from the Herpes Simplex Virus protein VP16). These constructs provide a proven reporter system that is independent of endogenous factors in mammalian cells and, therefore, well suited for studying the performance of the EXT assay with precisely controlled levels of GV. Moreover, the Split-TEV system relies on the nuclear translocation of the GV transcription factor to link the protease reconstitution assay to a transcriptional readout ([Wehr, Laage et al. 2006](#)). The performance of the modular GV-responsive

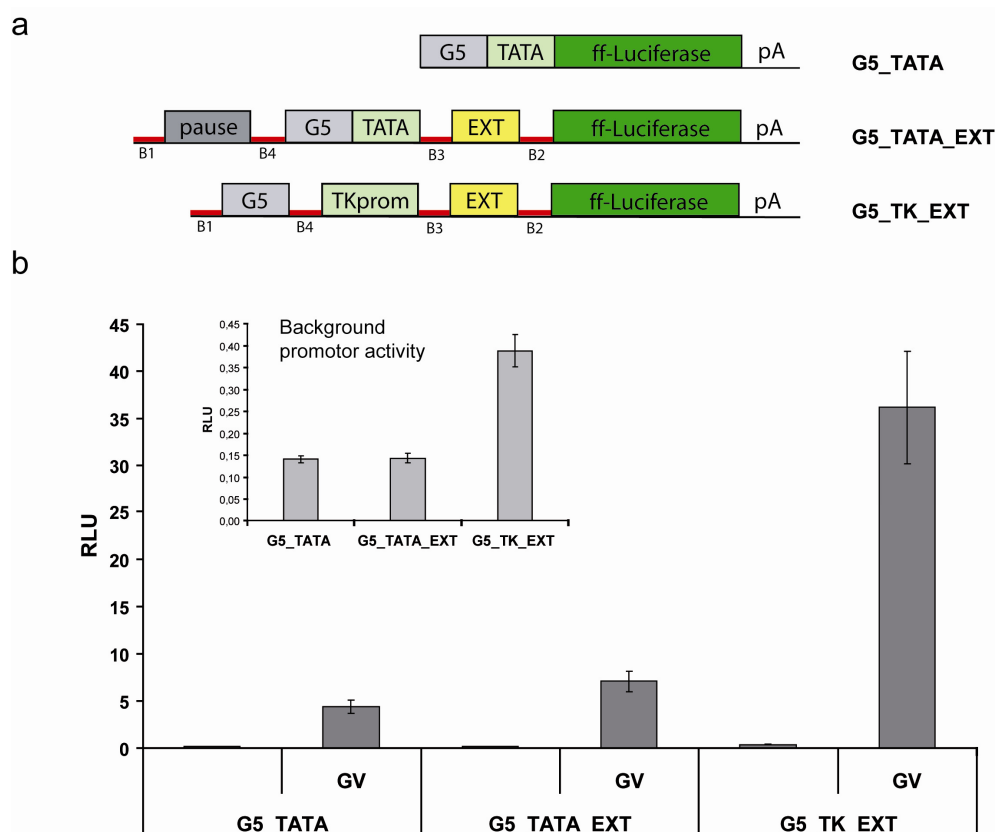


Figure 17. EXT reporter constructs. (a) Standard and modular reporter constructs with two different minimal promoters (TATA-box and TK promoter). “pause” stands for a transcriptional pause site and G5 for the five times Gal4 binding site. B1,B2,B3 and B4 designate recombination sites. **(b)** Functional test of the EXT reporter constructs by the Luciferase reporter gene assay. Three constructs depicted above were transfected into PC12 tetOFF cells with or without GV (Gal4-VP16 transcriptional activator). TATA-box EXT construct shows similar activation levels compared to the standard Luciferase reporter plasmid while the TK_EXT construct shows significantly higher signals in response to cotransfection of the GV. The inlay image shows the elevated background activity of the TK minimal promoter in comparison to the TATA-box.

constructs with both, TATA and TK minimal promoters (**Figure 17a**) was assessed by the luciferase reporter gene assay and compared to the performance of the standard G5-reporter carrying a TATA-box. Experiments with several modular EXT reporter plasmids have revealed that TATA-box carrying constructs perform as efficiently as the standard G5-luciferase reporter construct. They show equal background levels and similar induction rates upon cotransfection with GV. As expected, the stronger basal activity of the TK promoter displayed elevated levels of constitutive reporter expression and correspondingly higher maximal activation levels in PC-12 cells (**Figure 17**). The average induction rate (the maximal signal divided by the background) is similar for the TK promoter and for the TATA-box and lies between 100 and 1000 fold depending on the cell type and the time line of the experiment. Therefore both TK and TATA-box containing modular constructs provide a good measurement window for transcriptional activation.

Comparison of the "RNA" and luciferase reporter

We have performed a pilot experiment to get a first performance estimation of an RNA-based assay compared to a protein-based reporter gene assay. In the present design of the EXT assay reporter constructs, the EXT is expressed on the same mRNA as the firefly luciferase open reading frame. Considering the synthetic nature and the small size of the EXT, we assumed that the mRNA stability should be similar to the stability of the firefly luciferase mRNA alone. To assess the protein and mRNA stability, we used the tet-Off reporter system that allows the immediate shutting down of the reporter gene expression upon addition of doxycyclin to the medium ([Gossen and Bujard 2002](#)). The assay was performed in PC-12 tet-OFF cells stably transfected with the tTA transcriptional activator that binds to the tetracycline operator sequence (tetO) and drives constitutive expression of the reporter gene. Doxycyclin (Dox) sequesters the tTA protein, prevents it from binding to the DNA and inhibits transcription of the tetO reporter construct (**Figure 18a**). The cells transfected with the tetO controlled Luciferase reporter plasmid (pUHC13-3) were allowed to express the firefly Luciferase for 12 hours. The transcription was inhibited by adding Dox (500 ng/ml) and samples were collected at different time points. The abundance of the reporter protein was quantified as the Luciferase activity and the mRNA abundance by quantitative PCR relative to beta-actin mRNA. Control populations of cells received Dox simultaneously with the transfection and were used to determine the background expression levels of the reporter. Luciferase activity reached background levels after 48 hours (estimated half life ~6h) due to high stability of the firefly Luciferase protein (**Figure 18b**). In contrary, the Luciferase mRNA is depleted to background levels already at 8 hours after transcription has been shut down (estimated half life <2h) (**Figure 18c**). The tetO-luciferase construct used in this experiment carries an early simian virus 40 polyadenylation signal whereas the EXT reporter constructs embedded in the pGL3 reporter vector backbone are linked to the late SV40 polyA. This difference as well as the presence of the EXT might influence the RNA stability. The actual RNA stability for the modular EXT reporter constructs remains to be quantified but from this analysis it is clear that RNA reporter provide obviously a better kinetic resolution than a luciferase reporter.

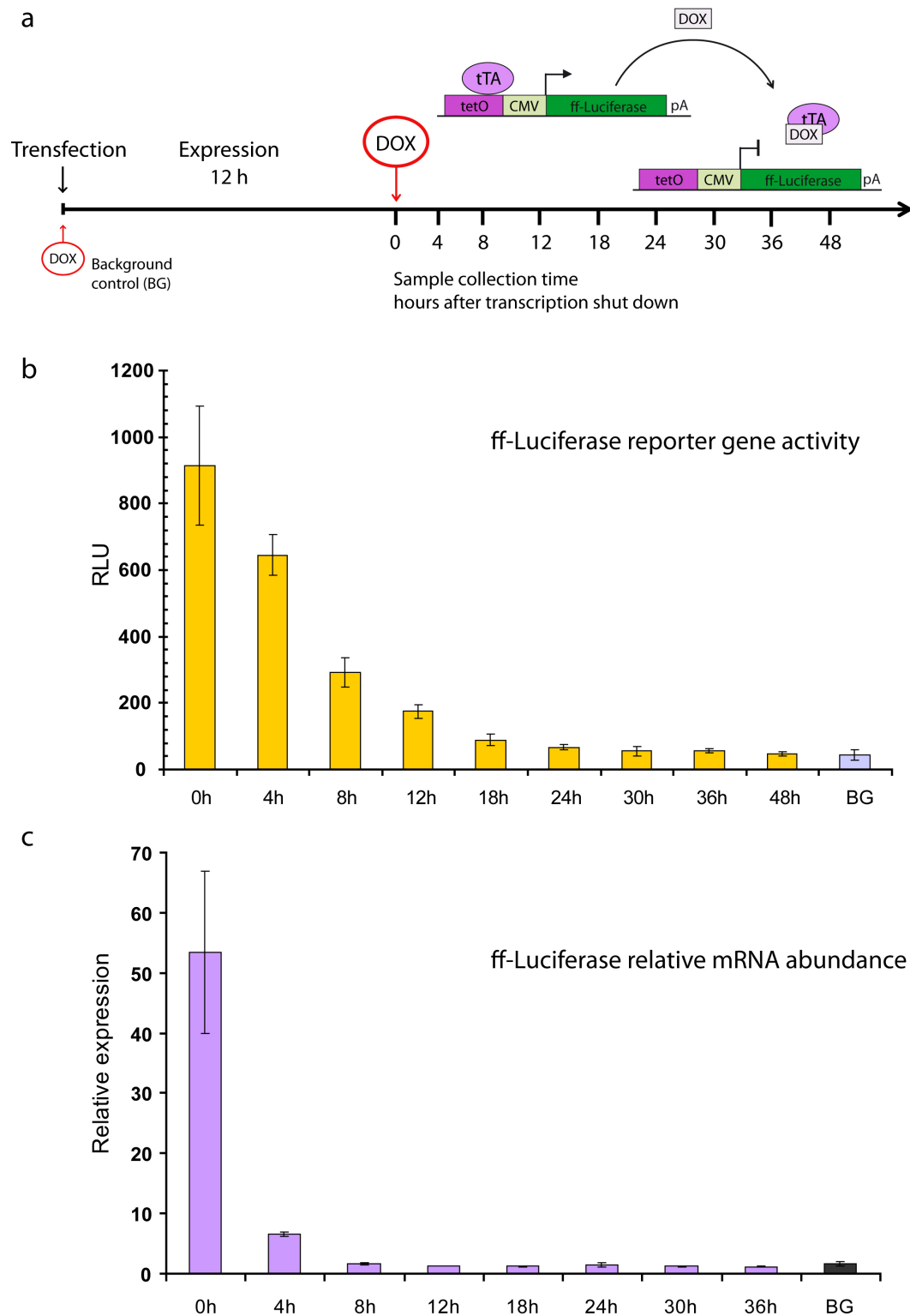


Figure 18. Stability of the Protein-based versus mRNA-based reporter. (a) PC-12 tetOFF cells were transfected with the tetO_CMV_Luciferase reporter constructs carrying the CMV promoter and the tetracycline responsive *cis*-element (tetO) that is activated by the tTA transcription factor. After 12 hours of Luciferase expression, the activity of tTA was inhibited by addition of Doxycyclin and the transcription was blocked. A control sample was exposed to doxycyclin from the beginning of the experiment to indicate the background level of the system. After transcription has ceased, the stability of Luciferase mRNA and protein was determined by taking samples at various time points. (b) The activity of the Luciferase persists for many hours and reaches the background at 48 hours after Doxycyclin addition. (c) The mRNA abundance of the Luciferase relative to b-actin mRNA declines to background levels already after 8 hours.

Multiplexed EXT analysis with microarrays

The first multiplexed EXT assay measurement with microarrays was performed in a model experiment using G5-TK-EXT-Luciferase reporter constructs. Transcriptional activation of the reporters was controlled by coexpression of GV under the control of the CMV promoter. To explore the sensitivity of the system, the dynamic signal was simulated by using different amounts of GV-encoding plasmid DNA from sample to sample. The experiment was controlled in parallel by the Luciferase activity measurement.

Because high numbers of “unregulated” signals are required for proper microarray normalization, we have allotted a large subset of EXT reporters (EXT₁₁₋₇₀) as a “reference EXT pool” and a smaller number (EXT₁₋₁₀) as the “assay pool” that served for the measurement of the dynamic signals. In total, 80 different EXT reporters were used in the analysis. The experiment was done in PC12 tet-OFF cells and five separate transfections were performed. Four subpopulations of cells received the 10 assay EXTs and corresponding increasing amounts of GV: 0ng, 0,1ng, 1ng or 10ng, respectively. The fifth and larger subpopulation of cells was transfected with the 60 EXT plasmids of the reference pool and a constant amount of GV. The DNA amount in the reference pool transfection was adjusted to elicit reasonably high and robust expression in the range of the expected assay pool signals. Several hours after transfection the cells from the assay pools were mixed with equal amounts of the reference pool, plated out and cultured under identical experimental conditions (**Figure 19a**). In this way, the reference pool EXTs were always processed together with the assay EXTs and could serve as an internal control for the expression, amplification and handling procedures.

The RNA and DNA were extracted from the cells 48 hours after transfection (**Figure 19a**). The RNA samples containing expressed EXT reporters served for the actual assay readout, whereas the DNA samples reflected the composition of the input plasmid DNA that was successfully transfected into the cells and actually contributed to the assay. We used the DNA input samples as an additional internal normalization control correcting for the pipetting or DNA concentration measurement errors and possible inhomogeneity in the transfection efficiency. For the RNA samples, the cDNA was prepared using random nanomer primers and the EXT reporters from both cDNA and DNA-input samples were

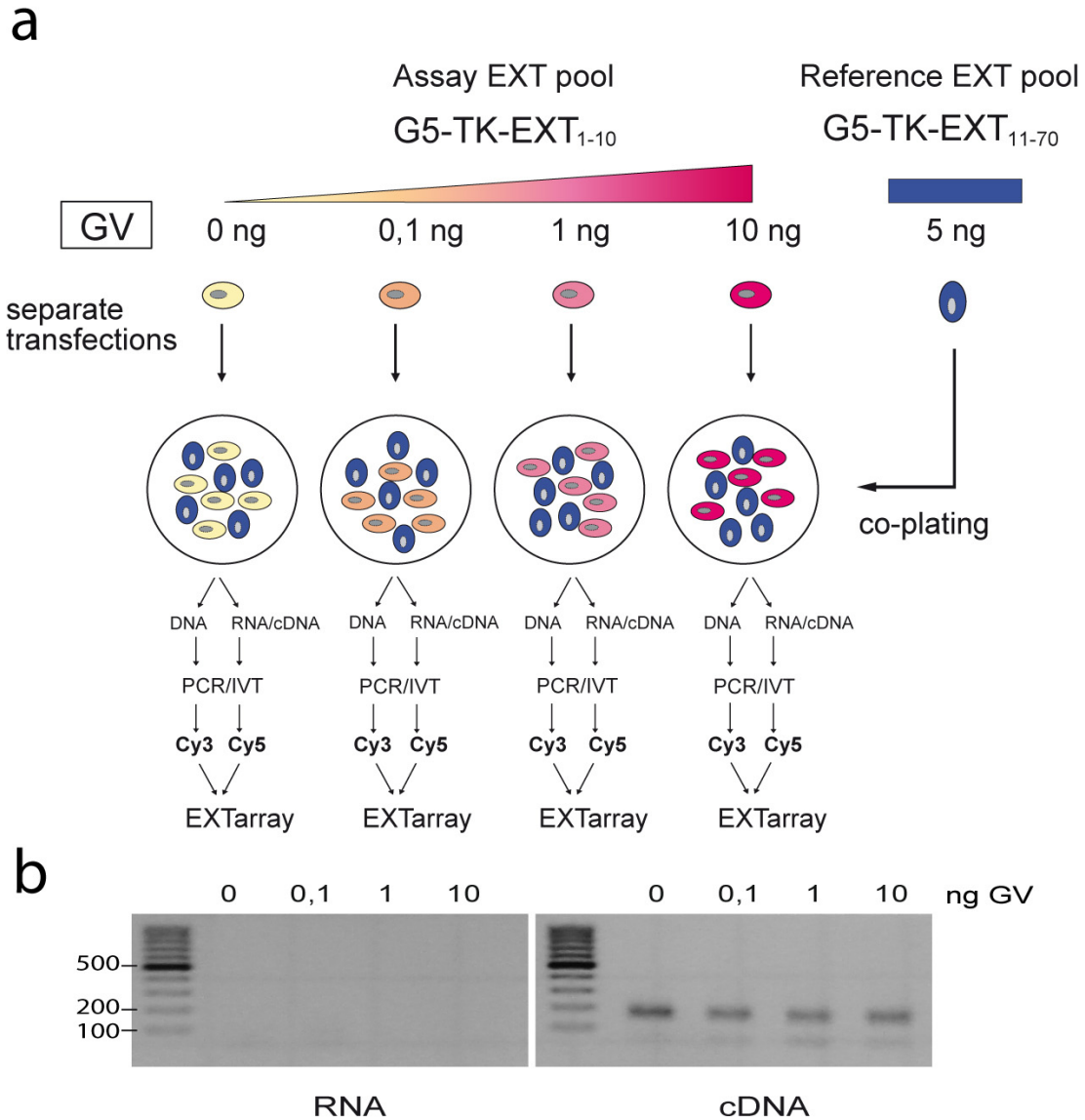
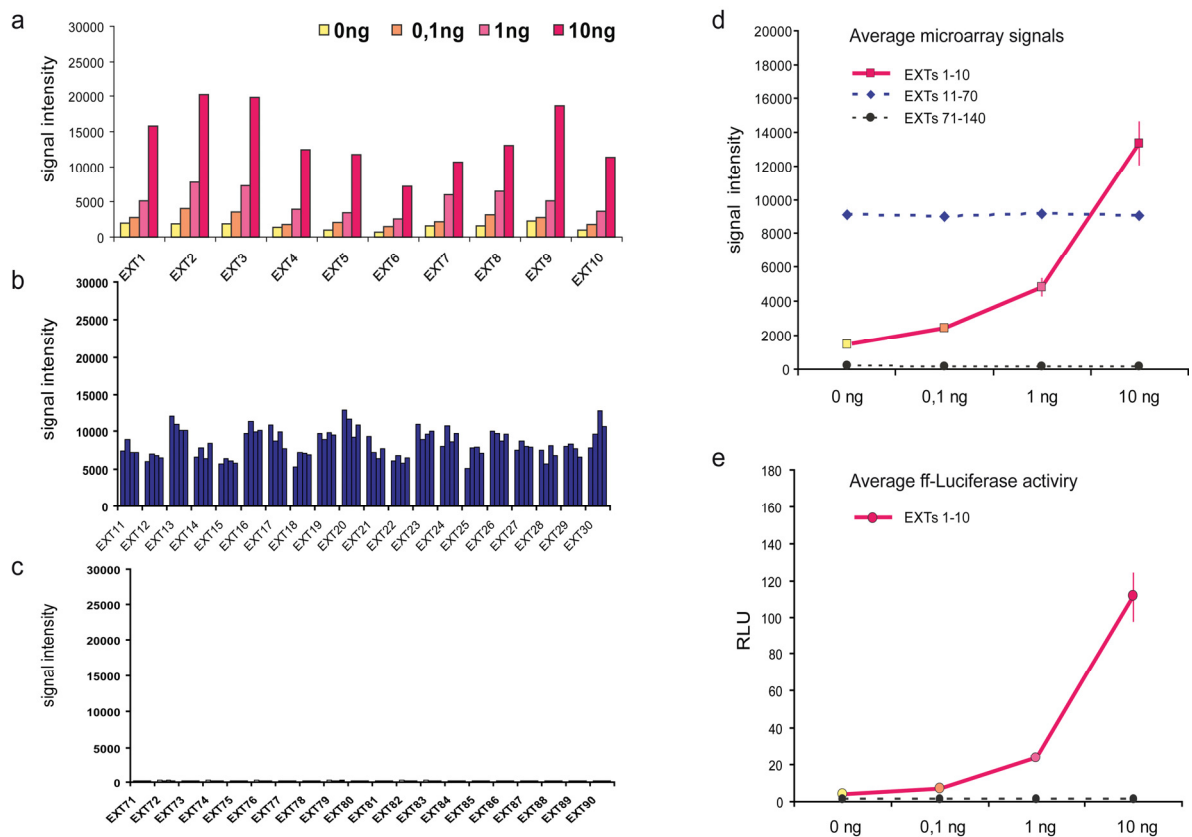


Figure 19. Testing EXT assay performance under control conditions. (a) The experimental design of the GV/G5 controlled multiplexed EXT assay. Increasing amounts of GV cotransfected with the set of G5-TK-EXT₁₋₁₀-Luciferase reporter plasmids (“assay EXT pool”) simulate a dynamic cellular signal. Equal number of cells transfected with the “reference EXT pool” (G5-TK-EXT₁₁₋₈₀-Luciferase) and a moderately high amount of GV are plated with each of the assay cell populations and serve as an internal control. After 48h at 37°C, RNA and DNA were isolated from all samples. The RNA was reverse transcribed into cDNA containing expressed EXT reporters and the DNA contained EXT encoding plasmids that entered the cells during transfection. The EXTs from both cDNA and plasmid DNA input were PCR amplified labeled with Cy5 and Cy3 dyes, respectively, and hybridized to a separate microarray for each of the 4 conditions. **(b)** Analysis of corresponding EXT PCR products obtained from cDNA samples with Dec1/Dec2 primers at 30 cycles. The negative control indicates that plasmid DNA contamination potentially present in the cDNA samples is not detectable at this cycle number.

PCR amplified for 30 cycles with the Dec1/Dec2 primer pair (**Figure 19b**). The PCR with the corresponding RNA samples controls for the possible plasmid DNA contamination in the RNA preparation. The 153 bp long PCR products were further used for in vitro transcription from T7 promoter to generate the 111b long RNA target for the microarray hybridization. The workflow of the EXT amplification is described in detail in materials and methods and illustrated in the **Figure 21**. DNA derived samples were labeled with Cy3 fluorescent dyes and the RNA-cDNA derived samples with Cy5. Each of the four samples that were transfected with the different GV amount was subsequently hybridized to a separate diagnostic microarray. Control EXTs c1-c5 were used in each hybridization for monitoring of the hybridization stringency. The cDNA derived fluorescent signals in the Cy5 channel were normalized to the plasmid DNA input signals in Cy3. The normalized signal intensities of the EXTs 1-10 increased in a dose-dependent manner with highly similar slopes (**Figure 20a**). In contrast, the signal intensities of EXTs 11-30 (derived from the reference EXT pool and exposed to a constant amount of GV) remained highly similar between the arrays (**Figure 20b**). The EXTs 71-90 that were not present in the transfection reflect overall microarray background and show virtually no signals which also indicates the specificity of hybridization (**Figure 20c**). The dose response curve of the microarray measurement (**Figure 20d**) appears very consistent with the supporting data from the luciferase reporter gene assay performed on the same samples (**Figure 20e**) The luciferase readings are derived from the EXT1-10 transfected cells that were isolated from the assay pool before the reference pool was added and plated onto 96-well plates with several replicate wells for each GV concentration. The microarray and the luciferase signals show very similar slopes, however, the Luciferase assay shows higher dynamic range (40-fold maximal induction vs. 8 fold for the microarray) which is likely due to the stability of the luciferase protein and enzymatic amplification of the signal.

From this analysis, we concluded that multiplexing reporter gene assays with a high number of EXTs is technically possible and reliably reflects expected dose-response effects.



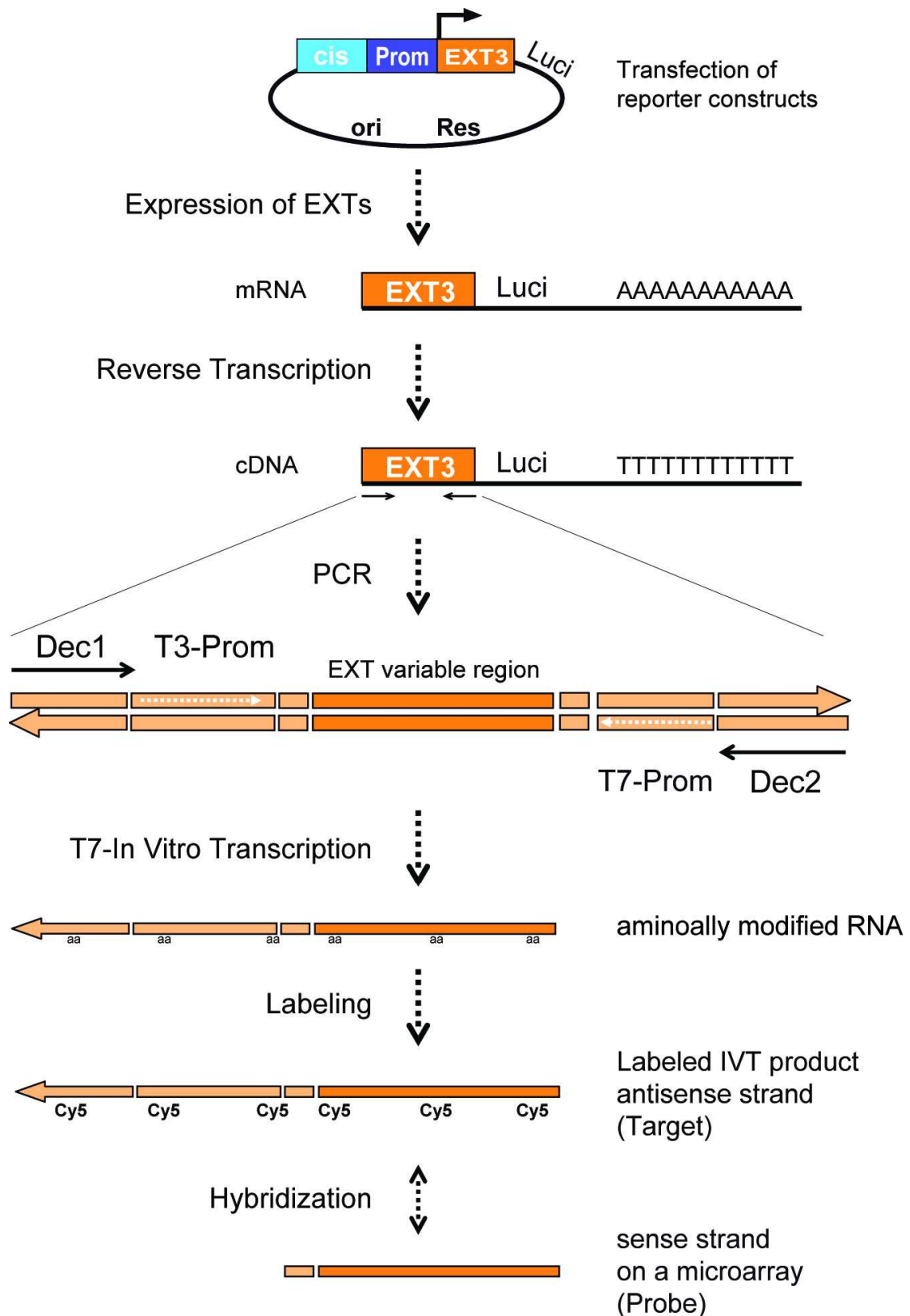


Figure 21. Experimental steps of the EXT analysis from transfections to microarray hybridization. The EXT mRNA is isolated from transfected cell populations, reverse transcribed and amplified by PCR. To generate single stranded target for microarray hybridization, in vitro transcription (IVT) is performed in the presence of aminoallyl-modified UTP. The RNA is then labeled by coupling of the cyanine dye esters (Cy3 or Cy5) to the aminoallyl moieties and the labeled product is hybridized to a microarray.

Monitoring of ErbB receptor activation and downstream signaling

We have applied the EXTassay approach to measure the activation of the ErbB receptor signaling pathway in response to Nrg1 (neuregulin 1). To demonstrate the versatility of the EXTassay, we decided to use the recently published split-TEV system to monitor selected protein-protein interactions involved in the initial steps of the ErbB signaling and combine it with the analysis of multiple *cis*-regulatory elements at the a more distant signaling level. We have used our library of 90 *cis*-regulatory EXT reporter constructs for the analysis of 30 different *cis*-regulatory elements (**Table 1**) with three unique EXTs each. We have measured ligand induced receptor dimerization and phosphorylation dependent recruitment of the downstream signaling proteins: SHC, Grb-2 and PI3Kp85α ([Wehr, Laage et al. 2006](#); [Wehr, Reinecke et al. 2008](#)). The design of the split-TEV assays is illustrated in **Figure 22**. Upon interaction, restored TEV protease activity leads to release of the Gal4-VP16 (GV) transcription factor.

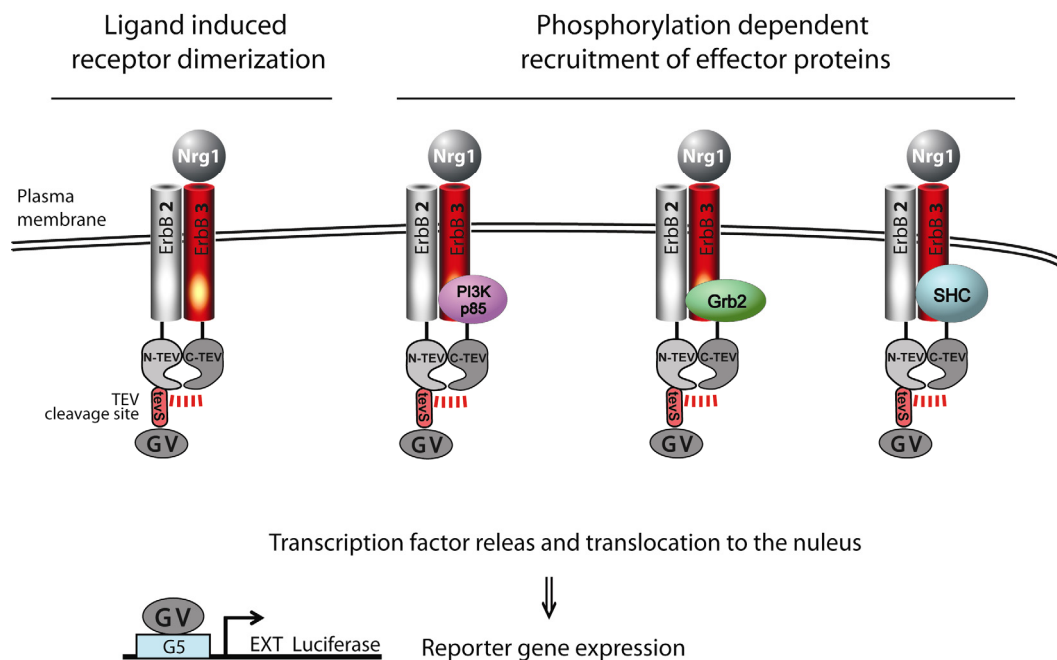
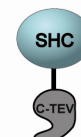
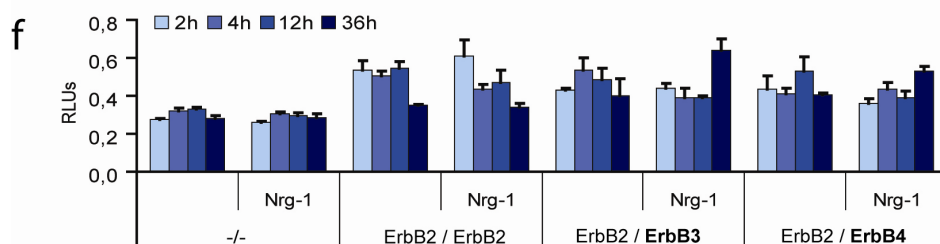
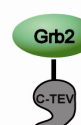
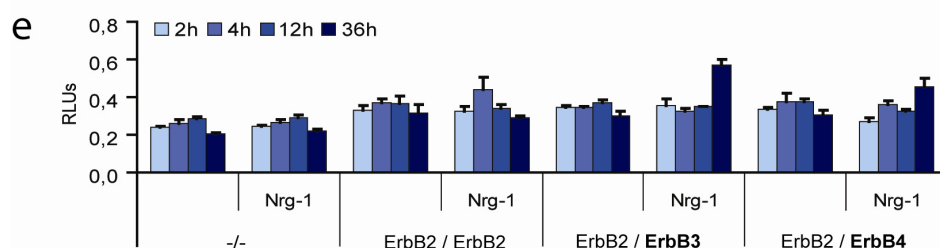
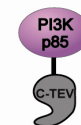
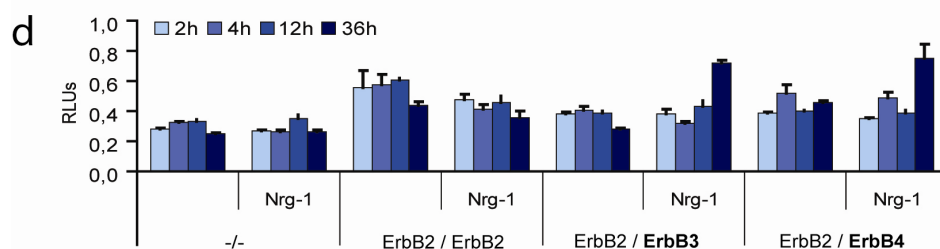
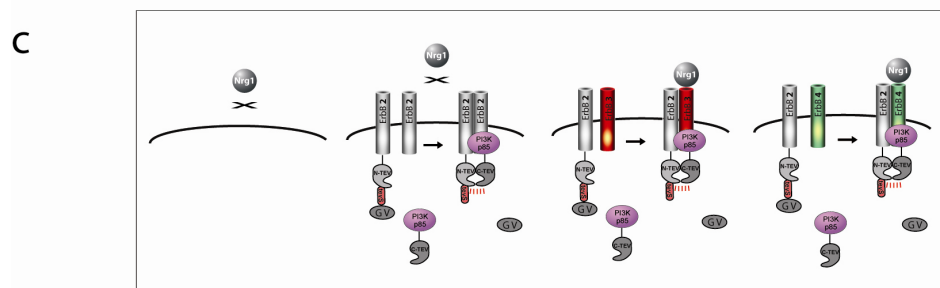
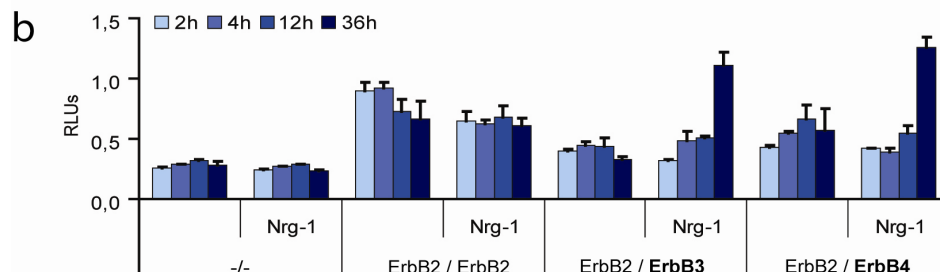
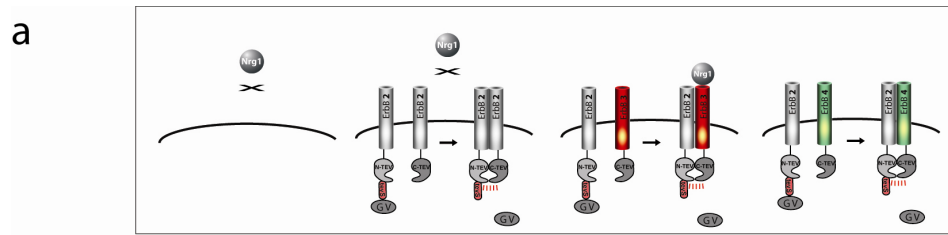


Figure 22. Split-TEV assays integrated into the analysis of the ErbB signaling pathway with EXTassays. For each receptor pair, four protein-protein interaction (PPI) assays were performed in parallel to the analysis of 30 *cis*-regulatory elements. An example is shown for the ErbB2/3 receptor pair (ErbB2/2 and ErbB2/4 were analyzed in the same way). The ErbB2 receptor fusion construct (ErbB2-N-TEV-GV) carries the N-terminal fragment of the TEV protease fused to the transcription factor (GV) via a TEV specific cleavage site (tevS). The other interaction partner carries the C-terminal fragment of the TEV protease. Upon interaction, restored protease activity leads to the release of the transcription factor GV that activates the expression of the responsive reporter gene (G5_EXT_Luciferase).

We used the GV-responsive G5_EXT_Luciferase reporter library to monitor the Split-TEV assays incorporated into the EXTassay format. Eight unique EXTs were assigned to each interaction. At the same time, single Luciferase activity measurements were performed from the same transfections and compared to the EXTassay data. The constructs encoding assay components for each protein-protein interaction were combined in a separate transfection with 8 unique EXT reporters carrying either TATA-box or TK promoter.

Out of four ErbB receptors only ErbB3 and ErbB4 can bind neuregulin. ErbB2 is deficient in ligand binding but acts as the preferred coreceptor for other ErbB proteins ([Graus-Porta, Beerli et al. 1997](#)). Therefore, we analysed two receptor heterodimers: ErbB2/ErbB3 and ErbB2/ErbB4, plus the ErbB2/ErbB2 homodimer and the control condition without any receptors (-/-).

For each receptor combination, five separate transfections were performed: one with 90 *cis*-element reporter constructs and four with corresponding split-TEV assays to ensure unambiguous coupling of EXT reporters. Four hours after transfection, the cells were mixed and plated out. We allowed 24 hours for expression of the receptors and subsequently stimulated the cells with soluble Nrg1 (**Figure 24a**) (NrgI- β 1a isoform EGF domain, courtesy of Dr. Tobias Fischer). We harvested the cells 2h, 4h, 12h and 36h after stimulation for luciferase assays and EXTassays. First, we monitored ErbB receptor dimerization and adapter recruitment by standard luciferase-coupled Split-TEV assays. Only ligand binding competent ErbB receptor complexes ErbB2/ErbB3 and ErbB2/ErbB4 responded to Nrg1 stimulation, detectable after 36 hours (**Figure 23**) in accordance with the literature and previous split-TEV experiments ([Wehr et al., 2006 and 2008](#)). In contrast, ErbB2/ErbB2 homodimers did not mediate Nrg1 effects but displayed elevated constitutive levels of dimerization over the negative control, in agreement with the previous results and the literature (**Figure 23b**). ErbB2 is known to elicit signaling upon overexpression in the absence of the ligand and can lead to the transformation of cells in culture ([Di Fiore, Pierce et al. 1987](#))



Receptor dimerization

Receptor-effector protein interactions

Figure 23. Luciferase measurements of the Split-TEV assays to monitor ErbB receptor interactions. (a) Schematic representation of the split-TEV assays for ErbB receptor activation measurements. Corresponding Luciferase data are shown below (b). (c) Schematic representation of the Split-TEV assays for the recruitment of the PI3K p85 α to the phosphorylated receptor dimer. The luciferase activity measurements are shown below (d). Corresponding recruitment assays for SHC and Grb-2 are depicted in (e) and (f). Signals visible at the (-/-) condition originate from the background activity of the TK promoter. Elevated signals displayed by the ErbB2/ErbB2 homodimer reflect ligand independent ErbB2 signaling triggered by overexpression. ErbB3 and ErbB4 receptors exhibit maximum 2-3 fold induction in receptor dimerization in response to Nrg1 stimulation expected for soluble Nrg1. The induction is not apparent at earlier time points (2h, 4h and 12h) as this time is probably too short for detectable Luciferase expression levels.

Next, we isolated plasmid DNA and RNA from the Nrg1 treated samples for microarray analysis as described for the previous experiment (see **Figure 19a**). A reference pool of 60 unique EXTs was added to each sample in roughly 5 fold excess before PCR amplification to ensure proper microarray normalization and balanced PCR amplification. Signals produced by expressed EXT reporters were normalized to the DNA input and to the average signal of the reference EXT pool. Correspondingly normalized EXT-reporter signals were plotted for the four time points (2h, 4h, 12h and 36h) and the four receptor conditions (no receptor (-/-), ErbB2/ErbB2 (2/2), ErbB2/ErbB3 (2/3) and ErbB2/ErbB4 (2/4)) (**Figure 24**). Note, that for clarity reasons not all individual *cis*-assays are plotted.

Already 2h after Nrg1 stimulation, we observed marked changes in the assay profiles dependent on the different ErbB receptor combinations (**Figure 24**). Most prominently, 2/2 and 2/4, but not 2/3, significantly suppressed “AP-1” *cis*-reporter activities 2-3 fold (2/2: $p < 0,005$; 2/4: $p < 0,01$). Receptor combinations 2/2, 2/4 and to a lesser extent 2/3 also reduced “p53” mediated transcription more than 3-fold (2/2: $p < 0,005$; 2/3: $p < 0,05$; 2/4: $p < 0,01$). In contrast, “SRE” reporter activity was elevated 4-fold by 2/3 ($p > 0,005$) whereas 2/2 and 2/4 had no effect at this time point.

At 4h, the overall reporter pattern remained highly similar compared to the 2h sample. Nonetheless, the “AP1” *cis*-reporter showed a tendency of activation in the ErbB 2/3 sample ($p = 0,07$ versus -/- at 4h, and $p = 0,09$ versus 2/3 at 2h)

The highest overall activity of the Nrg1 signaling was detected at 12 hours after stimulation and the pattern was considerably altered compared to the 2 and 4 hours time points. The pattern of the *cis*-element activation was characterized by the remarkably high activation AP1 (activator protein1) and SRE (serum response element).

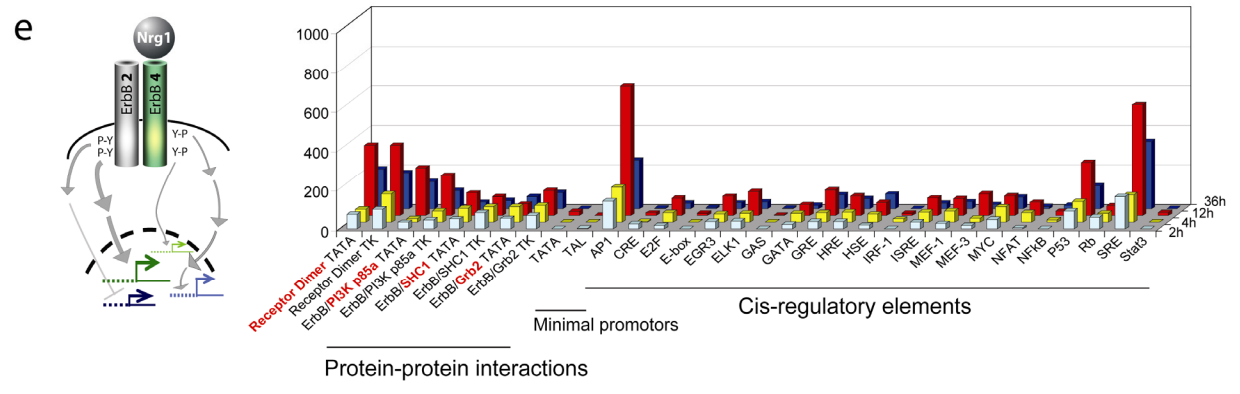
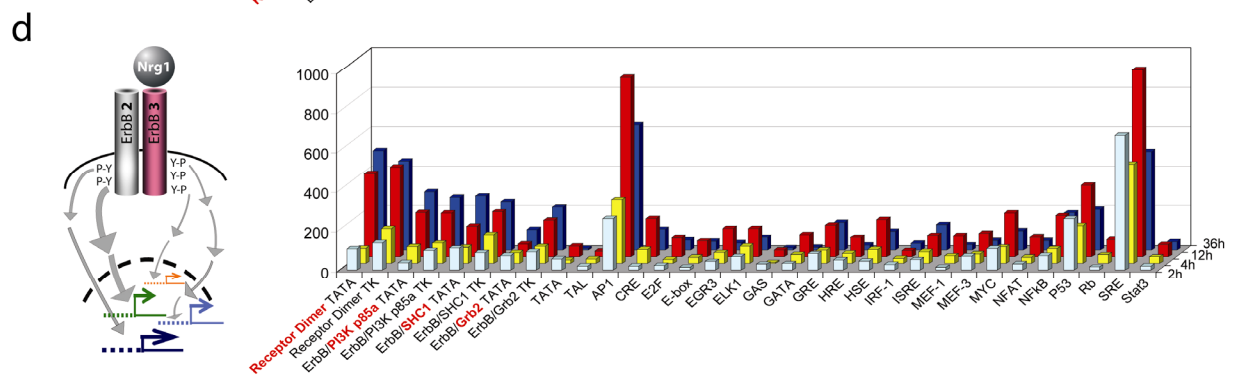
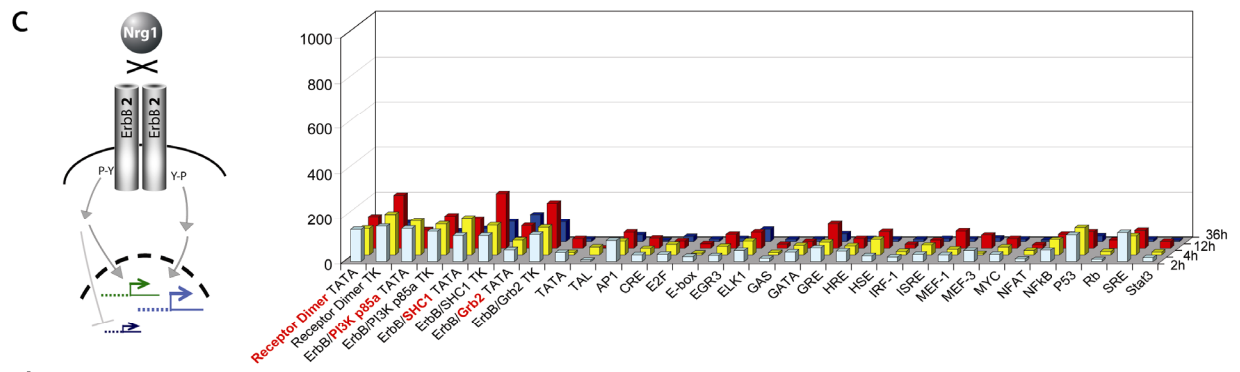
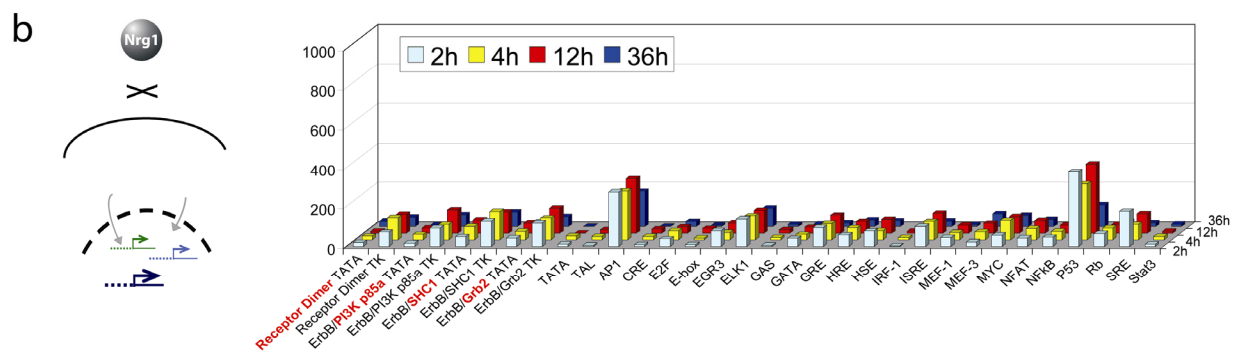
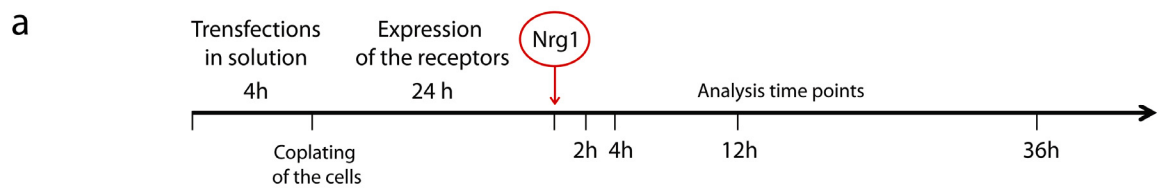


Figure 24. Kinetic measurement of the Nrg1/ErbB receptor signaling with EXTassays.

(a) PC-12 cells were transfected in solution with different assay components and cultured under identical experimental conditions for 24h to allow receptor expression. The cells were exposed to soluble Nrg1 and the measurements were performed at 2, 4, 12 and 36 hours after stimulation. Signals corresponding to different protein-protein interaction and *cis*-element assays are plotted along the X-axis. For each assay, average microarray intensities of several EXT reporters are shown. Measurements performed at different time points are plotted along the Z-axis. Figure (b) reflects endogenous signaling in PC-12 cells in the absence of ErbB receptors, (c) the overexpression of the Nrg1 independent ErbB2 protein. The results for Nrg1 induced signaling mediated by the ErbB2/3 and ErbB2/4 receptor complexes are shown in figures (d) and (e), respectively.

The activation patterns of the 2/3 and 2/4 receptor complexes were very similar at 12h after stimulation although marked differences with several *cis*-assays were detected. Most prominently, differential effects were observed for “CRE” and “NFkB” reporters that were elevated selectively in 2/3 but not in 2/4 or the control samples -/- and 2/2 (CRE: $p < 0,005$; NFkB: $p > 0,0005$). Therefore, receptor specific signaling mechanisms take place in PC-12 cells and can be detected with EXTassays. Differences between the Nrg1 binding competent ErbB complexes 2/3 and 2/4 became even more pronounced 36h after Nrg1 stimulation with ErbB2/3 receptor pairs being more effective both at the level of receptor activation and downstream signaling compared to ErbB2/4.

At 36 hours after stimulation (which corresponds to 60h post transfection) overall signal intensity declined probably due to signaling downregulation and possibly to the bias related to the transient transfection conditions. Reduction in signal intensities was observed also in the control samples -/- and 2/2, suggesting that this is independent of neuregulin. Nonetheless, ErbB 2/3 protein interaction assays and some *cis*-assays continue to show high signals indicating high potency particularly of this signaling complex.

Split-TEV assays monitoring dimerization and adapter recruitment revealed a strong Nrg1-dependent activation of ErbB2/3 and 2/4 receptor complexes. In contrast to the luciferase assay, ErbB protein-protein interactions could be clearly detected already at 12 hours demonstrating the advantage of the RNA based EXT reporter system over the protein based Luciferase reporter (**Figure 23, 24**). RNA based reporter allows faster measurement kinetics because protein synthesis and folding steps are omitted. Due to their lower basal activity, the TATA-box carrying EXT reporters showed better induction levels compared to the TK promoter driven EXTs. Nrg1 dependent ErbB2/3 and 2/4 receptor activation was robustly detected at the level of dimerization and recruitment of the PI3K p85 α subunit at 12h but not for SHC1 and Grb2.

Improved design of the EXT reporter constructs

In the course of experiments, we encountered a problem of DNA contamination in the RNA samples. None of the present RNA purification methods guarantees complete removal of genomic and especially plasmid DNA. For EXT assays, it is crucial to obtain DNA free samples. Carry over of plasmid DNA may influence the composition of complex EXT pools amplified by PCR and contribute to the background of the measurement. So far, we have solved the problem by careful monitoring the PCR amplification and avoiding high cycle numbers. To achieve optimal results without a risk of contaminating plasmid DNA, EXT reporter constructs were modified to carry an intron sequence adjacent to the EXT between the Dec1 and Dec2 primer binding sites. Intron containing plasmid DNA templates produced a long PCR product (800 bp) including both the EXT and the intron. On the other hand, after EXT expression the intron was removed by splicing and amplification on the cDNA template resulted in shorter PCR products of 150-200bp (**Figure 25a**) Based on the size difference, unambiguous separation of the plasmid DNA and cDNA derived PCR product are possible. We have generated a test construct carrying the rabbit beta globin intron (700 bp in size) between the TATA-box and the EXT. In front of the intron, a sense primer binding site (S1) was placed to serve for EXT amplification instead of the Dec1. PCR amplification on the plasmid DNA and cDNA template with S1/Dec2 primer pair has revealed that the intron was efficiently spliced after expression of the construct in PC-12 cells giving rise to appropriate 225bp PCR product instead of 793bp for the plasmid DNA template (**Figure 25b**). Even if both plasmid and cDNA templates were present in the PCR reaction, an amplification of the longer product was suppressed by competition with the more efficiently amplified short one. Hence, no physical separation of the two PCR products by gel extraction or other additional procedures seem to be required to obtain pure preparations of cDNA derived EXT reporters.

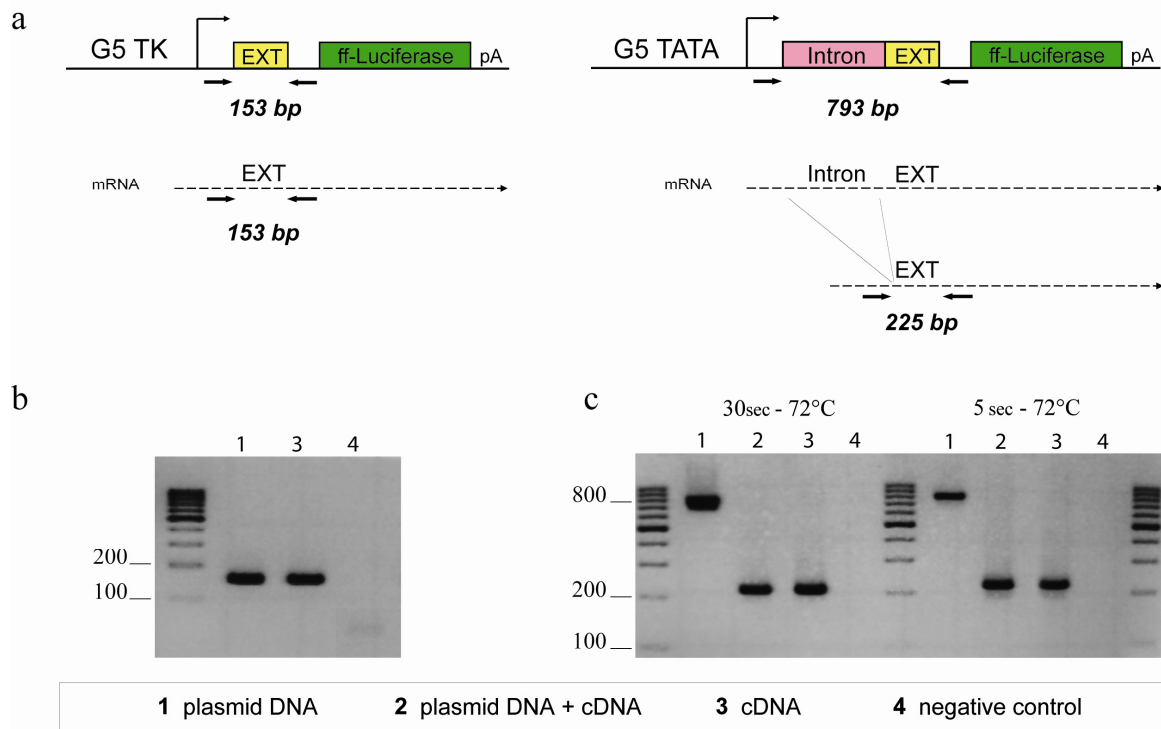


Figure 25. Improved design for EXT reporter constructs to exclude plasmid DNA contaminations. Trace amounts of plasmid DNA present in RNA preparations may contribute to the EXT amplification with Dec1/Dec2 primers producing a PCR product undistinguishable from that produced by the cDNA templates (**a**, **b**). PCR on a plasmid template carrying an intron sequence adjacent to the EXT generates a longer PCR products (800bp), whereas the cDNA derived PCR product is much shorter (225bp) due to the splicing of the intron (**a**, **c**). When both templates are present in the PCR reaction (sample2), the amplification of the longer product is suppressed with 5s as well as 30s extension times (**c**). In the case of the intron construct, the EXTs were amplified with S1/Dec2 primers for 30 cycles. Thus, the insertion of an intron sequence adjacent to the EXT allows specific EXT amplification from cDNA templates even in the presence of plasmid DNA contamination.

DISCUSSION

EXTassay readout characteristics

The probe design for gene expression analysis with microarray is substantially hindered by the limited choice of sequences that are available for particular targets. This gives rise to potential cross hybridization problems and imperfect fitting of hybridization temperatures of all probes. Synthetic “barcoding” DNA sequences as probes that were used in several recent studies still suffer, however to a lesser extent, from the same problems ([Silva, Marran et al. 2008](#)). As shown here by the series of control experiments, the synthetic EXT sequences employed by us behave in a highly predictable way avoiding any cross hybridization problems. In addition, an extensive panel of mismatch sequences allows an integrated and rigorous control over the hybridization quality of each array under investigation. Moreover, the issue of amplification induced bias has often not been addressed thoroughly before. Due to the homogeneous design of the short EXTs, amplification biases are excluded from our approach, as judged by strict control experiments. So far, barcoding strategies are mostly used in yeast for tracking of the mutant strains and in RNAi screens to determine cell survival but not to assay individual cellular components ([Giaever, Chu et al. 2002](#); [Silva, Marran et al. 2008](#)). It seems unlikely that the “unoptimized” barcodes used in these formats may allow similarly precise quantitative (rather than qualitative) cellular assays comparable to the EXTassays presented in this work. Nonetheless, we observed in EXTassays that EXT performed with a certain variability. The majority of the EXT reporters performed in the same range with slight variations although some “outliers” were detected. The variability of the EXT performance may originate from a variance in the expression level and/or measurement errors in the determination of plasmid DNA. We can, however, not formally exclude that the observed “outlier signals” may be attributed to principal EXT performance differences. Therefore, EXTassays in the current setup should be exclusively designed as comparative studies comparing identical EXTs with each other (similarly to gene expression experiments).

EXTassay and standard reporter gene assays

In its basic principle, EXTassay is a multiplexed reporter gene assay format and is not devoid of all the limitations associated with it. It is an indirect measurement system that relies on the cellular transcriptional machinery and can be influenced by general changes in transcription, pre-mRNA processing or RNA degradation. Each assay involves genetic modification of the cells with one or several DNA constructs. For more sophisticated assays, e.g. including protein-protein interactions, transient overexpression of modified proteins was employed. This caveat could be improved in the future by using other transfer systems (e.g. viral vectors) or by generating stable cell lines or batches. In most cellular assay formats (including EXTassays) proteins are tagged and their function might be influenced by the fused components. As a consequence, key findings with EXTassays should be independently verified. Another limitation is that EXTassay is not capable of providing fast “online” kinetic data compared to those obtained e.g. by fluorescence resonance energy transfer (FRET). Nevertheless, it allows following kinetic changes in the range of a few hours. The kinetic delay of the RNA based EXTassay measurement is already substantially reduced in comparison with the standard luciferase based reporter gene assays. EXT reporters do not require time for protein translation and folding and can not be influenced by mechanisms of translation regulation. Another kinetic advantage of the EXT over the firefly luciferase reporter concerns the signal persistence and the possibility to measure deactivation. Once expressed, luciferase protein is stable in the cell with a half life of app. 6h whereas the estimated half life of the EXT reporter mRNA is below 2 hours (**Figure 18**). Probably due to their ability to amplify signals, enzymatic reporter gene assays have been shown to be less reliable in quantitative measurements compared to RNA based techniques in the context of GPCR signaling ([Kovacs and Kaplan 1992](#)). For example, partial and full agonists might elicit equal luciferase reporter activities in GPCR activation assays ([Kemp, George et al. 1999](#)). Therefore, RNA-based reporters might be better suited for a quantitative monitoring of transient events involved in cellular signaling. The RNA stability may be further adjusted e.g. by introducing RNA destabilizing elements into the EXT reporter constructs.

The dynamic range of EXTassays is largely dependent on its readout system. Microarray hybridization provides a maximal signal range of three to five orders of

magnitude disregarding the background of the experiment related e.g. to promoter leakiness (depending on the platform). The realistic robust dynamic range of the microarray based EXTassay is, therefore 2 to 3 orders of magnitude. The dynamic range of the luciferase measurement can be much higher, as observed in the GV titration experiment with GV-mediated activation of the G5_TK_EXT_Luciferase reporters. However, this experiment involved rather high amounts of the potent synthetic transcription factor (Gal4-VP16), a situation hardly encountered in the context of real measurements. Quite the opposite, in the ErbB receptor activation experiment the Split-TEV luciferase measurement shows lower induction than corresponding EXT reporters. The measurement of the GV-titration was performed 48 hours after transfection when the RNA levels might have already declined while the ErbB activation was measured much earlier (at 2, 4, 12 and 36 hours). The kinetic advantage of the EXT reporter is also illustrated by the fact that the Nrg1 induced ErbB receptor dimerization was detected by the luciferase assay at 36 hours after stimulation but not at 12 hours (**Figure 23**). In contrast the EXT reporters showed maximal activation at 12 hour time point and were already reduced at 36 hours (**Figure 24**). In the future microarray hybridization might be replaced by high throughput RNA sequencing approaches for identification and quantification of the EXT reporters. The dynamic range of such readouts will certainly more accurately reflect real numbers.

The relevance and scalability of the EXTassay

With roughly 25000 protein encoding genes in the human genome and an average 2,5 splice variants per gene, the proteome size is estimated to be at least ~65000. The number of potential pairwise interactions formed by such a network is more than 2×10^9 . Another level of complexity is added by the posttranslational modifications that are known to significantly alter the protein functions and their interactions ([Papin, Hunter et al. 2005](#)).

Challenged by these numbers new technologies are moving towards development of highly multiplexed approaches. It is not likely that comprehensive analysis of the human proteome would be possible in the nearest future but it is important to create a foundation for further development of such techniques. Subsets of the proteome or

smaller networks could be analyzed already using the available technologies. For example, the number of transcription factors encoded in the human genome is estimated to ~1850 with 2×10^6 possible pairwise interactions (Papin, Hunter et al. 2005). Thus, the high number of unique EXT reporters that can be integrated in our system might be required if “systems-level” approaches may be addressed in the future.

In the proof of principle experiment we have used roughly 200 unique EXT reporter sequences quantified simultaneously by microarray hybridization. The maximal complexity of the EXT oligonucleotide library is $5,5 \times 10^{11}$. After extensive control experiments, we have shown that the minimal number of non-interfering EXT sequences that can be discriminated by microarray hybridization is at least 2×10^7 . The realistic number is definitely higher and should be determined with additional computational analysis. Therefore, the current design of the EXT reporter library allows simultaneous measurement of at least 10^7 (10 million) different cellular assays. The complexity of the measurement depends not only on the availability of the EXT reporters but also on the feasibility of the measurement by a microarray hybridization. The capacity of the modern microarray platforms reaches 125 000 probes per subarray in a 12-plex format for Nimblegen and 500 000 probes per subarray in a 96-plex format for the Affymetrix platform. The evaluation of the assays by using next generation sequencing principally provides even higher throughputs.

Flexibility of the EXT assay

The major advantage of the EXT assay is the flexibility with which measurements of very different cellular events are possible within one technically rather simple experiment. Any assay that can be linked to a transcriptional readout can be used in the frame of EXT assays. The list of such assays includes standard reporter gene assays formats such as *cis*-regulatory element activity measurement, transcription factor activation and protein/DNA binding covered by one-hybrid assays ([Deplancke, Dupuy et al. 2004](#)). The range of amenable assays continues with all types of transcriptionally coupled protein complementation systems such as two hybrid assay ([Luo, Batalao et al. 1997](#)), Split-TEV ([Wehr, Laage et al. 2006](#)), split-intein ([Kanno,](#)

[Ozawa et al. 2006](#)) split-ubiquitin ([Stagljar, Korostensky et al. 1998](#)) and proximity based assays e.g. mammalian protein-protein interaction trap assay (MAPPIT) and its modifications ([Lemmens, Lievens et al. 2008](#)) and the full length TEV based Tango assay ([Barnea, Strapps et al. 2008](#)). The mentioned techniques allow measurement of various protein-protein interactions including inducible, transient and modification dependent ([Wehr, Reinecke et al. 2008](#)). Additionally, any other events affecting the proximity of cellular components can be measured with these approaches e.g. proteolytic cleavage, protein degradation, translocation or specific subcellular localizations.

A set of desired assays is assembled and encoded by coupling to the unique EXT reporters. Thereafter all the variety of assays can be performed in one large scale measurement by single microarray hybridizations.

The capacity to accommodate large assay numbers and the unique versatility of the EXTassay approach makes it a potentially suitable tool for high throughput analysis of cellular signaling that involves very different types of cellular events.

The main bottleneck on the way to comprehensive analysis of signaling networks with EXTassays is the generation and integration of new genetically encoded assays. Each assay is introduced to the cells in the form of one or several DNA constructs. Some assays are contained in a single DNA construct carrying e.g. a *cis*-regulatory element and a unique EXT reporter sequence. Such constructs can be easily generated and applied on a large scale. However, sequence verification is required to validate the appropriate coupling of the EXT. In the future, *in situ* oligonucleotide synthesis could be used to generate high numbers of *cis*-elements readily linked to a unique EXT sequence on one DNA stretch.

Other assays involve expression of modified proteins and are encoded by more than one plasmid. Within the proof of principle experiment, we have performed the measurement of four such assays with the help of the split-TEV system. For each assay, a separate transfection was performed using two constructs encoding respective interaction partner proteins cotransfected with an appropriate unique EXT reporter plasmid. The separate transfection in this case is unavoidable to ensure unambiguous coupling to a specific EXT. Cells carrying different assays were combined after transfection and later assayed simultaneously. However, multiple transfections are inconvenient and time consuming. An ideal assay should require only one transfection. As a possible solution all the assay components could be cloned into one

DNA construct with a defined reporter EXT using. Using viral DNA transfer the constructs could be delivered to the cells in the amount adjusted such that every cell would receive maximally one construct.

An easier solution towards multiplexed EXTassays could be performed by using the newly developed transfected cell arrays ([Bailey, Wu et al. 2002](#)). To apply this technique, appropriate combinations of plasmid DNA are spotted on the glass surface and overlaid with transfection reagents. The cells plated on top of such a surface receive defined combinations of plasmids different from spot to spot.

Every assay to be integrated into the multiplexed measurement has to be functionally verified with respect to its performance. This requirement is the main difficulty that delays the development of the comprehensive EXTassay setup. However, once established the assay can be easily applied to very different studies to elucidate signaling mechanisms under different cellular context. Moreover, new assays can be incorporated at any time to improve the scope of the analysis.

There are numerous potential applications of the EXTassay. It can be easily applied in the near future to quantitative profiling of the transcription factor activities for comprehensive analysis of the known and characterized transcription factors, the number of which in human genome is below 2000 ([Papin, Hunter et al. 2005](#)). If preferred, *cis*-element activity analysis can be combined with the standard gene expression microarray measurement to study the mechanisms of transcription regulation. Moreover, the EXT oligonucleotide library may be used as an improved barcoding strategy for the cellular survival assays ([Silva, Marran et al. 2008](#)). Even more importantly, one could apply the EXTassay to high throughput screening of small molecules. We hope that this assay will be useful for monitoring drug effects in living cells for example to identify the mechanisms of action for established drugs or to identify potential new therapeutically active substances at the network level.

ErbB signaling measurement with EXTassays

In a biologically relevant proof of principle study, we were able to quantitatively measure the Nrg1 mediated activation of ErbB receptor tyrosine kinases with multiple diverse assays in one measurement. Our measurements of protein-protein interactions associated with the stimulated ErbB receptor complexes revealed several observations that are consistent with already published data. For example, we have detected ligand independent homodimerization of the ErbB2 protein accompanied by changes in the activity of several *cis*-regulatory elements. This is in good agreement with the ability of ErbB2 to induce transformation in cell culture by eliciting ligand independent signaling ([Di Fiore, Pierce et al. 1987](#)). The highest activation levels of both, protein interaction assays and *cis*-regulatory elements was exhibited by the ErbB2/ErbB3 heterodimer expressing cells suggesting that this receptor combination demonstrates enhanced signaling properties, consistent with the literature ([Alimandi, Romano et al. 1995](#)).

ErbB proteins are the most extensively studied group of receptor tyrosine kinases. First order signaling complexes of each ErbB receptor were previously characterized with biochemical in vitro approaches ([Schulze, Deng et al. 2005](#)). These techniques, however, do not provide any direct information about the downstream signaling in a particular cellular context. By using EXTassays, we were for the first time able to link the activation of distinct ErbB receptor complexes to activation of certain *cis*-regulatory elements in PC-12 cells by a simple and parallel measurement. Among the interesting observations is the reduction in p53 and AP-1 *cis*-regulatory element activities in response to ErbB2 overexpression. We have also been able to measure differential downstream signaling induced by distinct ErbB receptor complexes. For example, the AP-1 *cis*-regulatory element was differentially regulated by distinct ErbB receptor complexes. In the case of ErbB2 overexpression the AP-1 activity was suppressed relative to the control condition, while ErbB2/4 and ErbB2/3 receptor complexes have induces its activation. The *cis*-elements CRE and NFkB were shown to be exclusively activated by ErbB2/3 but not by ErbB2/4.

The two *cis*-regulatory elements that exhibited the most prominent activation are SRE and AP-1, which are classical elements involved in growth factor signaling. SRE was primarily identified as the part of the c-fos promoter that mediated its response to the serum and is bound by the serum response factor (SRF) ([Treisman 1986](#)) AP-1 is an

element recognized by the protooncogenic transcription factors of the c-Jun family in the form of homodimers or heterodimers with the c-Fos and ATF proteins. Two major pathways triggered by the Nrg1 dependent ErbB receptor activation are mitogen activated protein kinase (MAPK) pathway and phosphatidylinositol 3-kinase (PI3K) pathway ([Mei and Xiong 2008](#)). In order to analyse the mechanism of *cis*-element activation in more detail, further experiments are definitely required. One could use pathway specific small molecule inhibitors or RNAi targeted against different components of the MAPK or PI3K pathways to determine which of these might participate in the induction of certain *cis*-regulatory elements. For large scale profiling of *cis*-element activation patterns in response to Nrg1, it might be useful to generate cell lines stably transfected with different receptor combinations. This would allow more reliable measurement free from the possible bias introduced by the transient transfection assays that were used so far.

CONCLUSIONS

In the present work, we have established an assay for the simultaneous and quantitative monitoring of several distinct cellular assays which we termed EXTassay. An “oligonucleotide” EXT-based reporter system was created that allows upscaling of the assay principally to high throughput levels. Therefore, we have developed a particular custom microarray design and optimized protocols for target preparation and hybridization. This allows a robust and specific identification of unique oligonucleotide reporters by microarray hybridization. The performance of the system was validated by numerous hybridization controls and a model experiment with synthetic transcription factor.

Moreover, we successfully performed a proof of principle experiment using signaling proteins of the ErbB receptor tyrosine kinase family to assess their immediate and downstream signaling in PC-12 cells. The experiment demonstrated that the EXTassay approach can be used for monitoring differential signaling events at various levels of signal transduction networks. Ligand induced protein-protein interactions of the ErbB receptors at the plasma membrane were measured along with multiple DNA-regulatory elements at the transcriptional level. The obtained data were in the line with the current concepts of the ErbB biology suggesting the validity of the EXTassay based measurements. In addition, these experiments provided for the first time comprehensive and quantitative data of ErbB receptor signaling at the level of activated transcription factors.

Curriculum Vitae

PERSONAL INFORMATION

Surname, Name Botvinnik Anna
Address Jüdenstr. 13a, App13, D-37073, Göttingen.
E-mail botvinnik@em.mpg.de,
abotvinnik@googlemail.com
Nationality Ukrainian
Date and place of birth 26 October 1979, Vinnitsa, Ukraine

EDUCATION AND TRAINING

- Since May 2005*** PhD student
International Max Planck Research School for Molecular
Biology
*Max Planck Institute for Experimental Medicine,
Research group "Gene expression", Göttingen, Germany.*
- Project title: "Combining induced protease fragment assembly and
microarray analysis to monitor signaling networks in living
cells"
- 2003-2005*** M.Sc. in Molecular Biology
International Max Planck Research School for Molecular
Biology,
*Max Planck Institute for Biophysical Chemistry
Department Membrane Biophysics, Göttingen, Germany.*
- Project title: "Probing protein-protein interactions in clathrin-
mediated endocytosis using XFP fusion constructs"
- 1998-2003*** Diploma in Biochemistry
M.V. Lomonosov Moscow State University,
Department of Bioorganic Chemistry, Russian Federation
Institute of Bioorganic Chemistry, Russian Academy of Sciences
- Project title: "Genetic variability of *M. tuberculosis* clinical isolates"
- 1998*** Secondary School Certificate, Vinnitsa, Ukraine

LIST OF PUBLICATIONS

Journal articles:

Botvinnik A, Wichert S, Fischer TM, Rossner MJ
Integrated analysis of ErbB receptor activation and downstream signaling with EXTassays. (in revision)

Wehr MC, Reinecke L, Botvinnik A, Rossner MJ.
Analysis of transient phosphorylation-dependent protein-protein interactions in living mammalian cells using split-TEV. *BMC Biotechnol.* 2008 Jul 13;8:55.

Azhikina T, Gvozdevsky N, Botvinnik A, Fushan A, Shemyakin I, Stepanshina V, Lipin M, Barry C 3rd, Sverdlov E. A genome-wide sequence-independent comparative analysis of insertion-deletion polymorphisms in multiple Mycobacterium tuberculosis strains. *Res Microbiol.* 2006 Apr;157(3):282-90. Epub 2005 Sep 15.

Conference abstracts:

Botvinnik A, Wehr MC, Fischer TM, Rossner MJ. Measurement of the neuregulin – ErbB receptor signaling using split-TEV. (2007) *International PhD Student Symposium, Horizons in Molecular Biology, Göttingen, Germany*

Botvinnik A, Wichert S, Rossner MJ
Towards Barcode Arrays for the Integrated Monitoring of Cellular Signaling. (2008) *3rd ESF Functional Genomics Conference. Innsbruck, Austria.*

Djannatian M, Fischer TM, Botvinnik A, Rossner MJ. Monitoring G-protein-coupled receptor activation and downstream signaling using Split-TEV. (2008) *6th Forum of European Neuroscience. Geneva. Switzerland.*

BIBLIOGRAPHY

- Agid, Y., G. Buzsaki, et al. (2007). How can drug discovery for psychiatric disorders be improved? *Nat Rev Drug Discov* 6(3): 189-201.
- Alimandi, M., A. Romano, et al. (1995). Cooperative signaling of ErbB3 and ErbB2 in neoplastic transformation and human mammary carcinomas. *Oncogene* 10(9): 1813-21.
- Araujo, R. P., L. A. Liotta, et al. (2007). Proteins, drug targets and the mechanisms they control: the simple truth about complex networks. *Nat Rev Drug Discov* 6(11): 871-80.
- Bailey, S. N., R. Z. Wu, et al. (2002). Applications of transfected cell microarrays in high-throughput drug discovery. *Drug Discov Today* 7(18 Suppl): S113-8.
- Barnea, G., W. Strapps, et al. (2008). The genetic design of signaling cascades to record receptor activation. *Proc Natl Acad Sci U S A* 105(1): 64-9.
- Birnboim, H. C. and J. Doly (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res* 7(6): 1513-23.
- Brenner, S., S. R. Williams, et al. (2000). In vitro cloning of complex mixtures of DNA on microbeads: physical separation of differentially expressed cDNAs. *Proc Natl Acad Sci U S A* 97(4): 1665-70.
- Chanda, S. K. and J. S. Caldwell (2003). Fulfilling the promise: drug discovery in the post-genomic era. *Drug Discov Today* 8(4): 168-74.
- Cheo, D. L., S. A. Titus, et al. (2004). Concerted assembly and cloning of multiple DNA segments using in vitro site-specific recombination: functional analysis of multi-segment expression clones. *Genome Res* 14(10B): 2111-20.
- Chomczynski, P. (1993). A reagent for the single-step simultaneous isolation of RNA, DNA and proteins from cell and tissue samples. *Biotechniques* 15(3): 532-4, 536-7.
- Chomczynski, P. and N. Sacchi (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal Biochem* 162(1): 156-9.
- Citri, A. and Y. Yarden (2006). EGF-ERBB signaling: towards the systems level. *Nat Rev Mol Cell Biol* 7(7): 505-16.
- Deplancke, B., D. Dupuy, et al. (2004). A gateway-compatible yeast one-hybrid system. *Genome Res* 14(10B): 2093-101.
- DeRisi, J. L., V. R. Iyer, et al. (1997). Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* 278(5338): 680-6.
- Di Fiore, P. P., J. H. Pierce, et al. (1987). erbB-2 is a potent oncogene when overexpressed in NIH/3T3 cells. *Science* 237(4811): 178-82.
- DuBridges, R. B., P. Tang, et al. (1987). Analysis of mutation in human cells by using an Epstein-Barr virus shuttle system. *Mol Cell Biol* 7(1): 379-87.
- Eyckerman, S., I. Lemmens, et al. (2002). Design and use of a mammalian protein-protein interaction trap (MAPPIT). *Sci STKE* 2002(162): PL18.
- Fishman, M. C. and J. A. Porter (2005). Pharmaceuticals: a new grammar for drug discovery. *Nature* 437(7058): 491-3.
- Fodor, S. P., R. P. Rava, et al. (1993). Multiplexed biochemical assays with biological chips. *Nature* 364(6437): 555-6.
- Friedman, A. and N. Perrimon (2007). Genetic screening for signal transduction in the era of network biology. *Cell* 128(2): 225-31.

- Garrett, T. P., N. M. McKern, et al. (2003). The crystal structure of a truncated ErbB2 ectodomain reveals an active conformation, poised to interact with other ErbB receptors. *Mol Cell* 11(2): 495-505.
- Gertz, J. and B. A. Cohen (2009). Environment-specific combinatorial cis-regulation in synthetic promoters. *Mol Syst Biol* 5: 244.
- Giaever, G., A. M. Chu, et al. (2002). Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* 418(6896): 387-91.
- Giaever, G., D. D. Shoemaker, et al. (1999). Genomic profiling of drug sensitivities via induced haploinsufficiency. *Nat Genet* 21(3): 278-83.
- Gossen, M. and H. Bujard (2002). Studying gene function in eukaryotes by conditional gene inactivation. *Annu Rev Genet* 36: 153-73.
- Graus-Porta, D., R. R. Beerli, et al. (1997). ErbB-2, the preferred heterodimerization partner of all ErbB receptors, is a mediator of lateral signaling. *EMBO J* 16(7): 1647-55.
- Greene, L. A. and A. S. Tischler (1976). Establishment of a noradrenergic clonal line of rat adrenal pheochromocytoma cells which respond to nerve growth factor. *Proc Natl Acad Sci U S A* 73(7): 2424-8.
- Hartley, J. L., G. F. Temple, et al. (2000). DNA cloning using in vitro site-specific recombination. *Genome Res* 10(11): 1788-95.
- Hill, S. J., J. G. Baker, et al. (2001). Reporter-gene systems for the study of G-protein-coupled receptors. *Curr Opin Pharmacol* 1(5): 526-32.
- Holbro, T. and N. E. Hynes (2004). ErbB receptors: directing key signaling networks throughout life. *Annu Rev Pharmacol Toxicol* 44: 195-217.
- Hopkins, A. L. (2008). Network pharmacology: the next paradigm in drug discovery. *Nat Chem Biol* 4(11): 682-90.
- Hynes, N. E. and H. A. Lane (2005). ERBB receptors and cancer: the complexity of targeted inhibitors. *Nat Rev Cancer* 5(5): 341-54.
- Johnston, P. A. (2002). Cellular platforms for HTS: three case studies. *Drug Discov Today* 7(6): 353-63.
- Jones, R. B., A. Gordus, et al. (2006). A quantitative protein interaction network for the ErbB receptors using protein microarrays. *Nature* 439(7073): 168-74.
- Kanagawa, T. (2003). Bias and artifacts in multitemplate polymerase chain reactions (PCR). *J Biosci Bioeng* 96(4): 317-23.
- Kanno, A., T. Ozawa, et al. (2006). Intein-mediated reporter gene assay for detecting protein-protein interactions in living mammalian cells. *Anal Chem* 78(2): 556-60.
- Kemp, D. M., S. E. George, et al. (1999). Partial agonism at serotonin 5-HT_{1B} and dopamine D_{2L} receptors using a luciferase reporter gene assay. *Eur J Pharmacol* 373(2-3): 215-22.
- Kholodenko, B. N., O. V. Demin, et al. (1999). Quantification of short term signaling by the epidermal growth factor receptor. *J Biol Chem* 274(42): 30169-81.
- Kibbe, W. A. (2007). OligoCalc: an online oligonucleotide properties calculator. *Nucleic Acids Res* 35(Web Server issue): W43-6.
- Kitano, H. (2007). A robustness-based approach to systems-oriented drug design. *Nat Rev Drug Discov* 6(3): 202-10.
- Kotarsky, K., N. E. Nilsson, et al. (2003). Progress in methodology. Improved reporter gene assays used to identify ligands acting on orphan seven-transmembrane receptors. *Pharmacol Toxicol* 93(6): 249-58.

- Kovacs, D. M. and B. B. Kaplan (1992). Discordant estimates of heterologous promoter activity as determined by reporter gene mRNA levels and enzyme activity. *Biochem Biophys Res Commun* 189(2): 912-8.
- Krutzik, P. O. and G. P. Nolan (2006). Fluorescent cell barcoding in flow cytometry allows high-throughput drug screening and signaling profiling. *Nat Methods* 3(5): 361-8.
- Lemmens, I., S. Lievens, et al. (2008). MAPPIT: a versatile tool to study cytokine receptor signaling. *Biochem Soc Trans* 36(Pt 6): 1448-51.
- Li, X., X. Jiang, et al. (2006). High throughput assays for analyzing transcription factors. *Assay Drug Dev Technol* 4(3): 333-41.
- Lum, P. Y., C. D. Armour, et al. (2004). Discovering modes of action for therapeutic compounds using a genome-wide screen of yeast heterozygotes. *Cell* 116(1): 121-37.
- Luo, Y., A. Batalao, et al. (1997). Mammalian two-hybrid system: a complementary approach to the yeast two-hybrid system. *Biotechniques* 22(2): 350-2.
- MacDonald, M. L., J. Lamerdin, et al. (2006). Identifying off-target effects and hidden phenotypes of drugs in human cells. *Nat Chem Biol* 2(6): 329-37.
- Mannocci, L., Y. Zhang, et al. (2008). High-throughput sequencing allows the identification of binding molecules isolated from DNA-encoded chemical libraries. *Proc Natl Acad Sci U S A* 105(46): 17670-5.
- Mei, L. and W. C. Xiong (2008). Neuregulin 1 in neural development, synaptic plasticity and schizophrenia. *Nat Rev Neurosci* 9(6): 437-52.
- Michnick, S. W. (2003). Protein fragment complementation strategies for biochemical network mapping. *Curr Opin Biotechnol* 14(6): 610-7.
- Mullis, K., F. Faloona, et al. (1986). Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harb Symp Quant Biol* 51 Pt 1: 263-73.
- Natarajan, M., K. M. Lin, et al. (2006). A global analysis of cross-talk in a mammalian cellular signaling network. *Nat Cell Biol* 8(6): 571-80.
- Nuwaysir, E. F., W. Huang, et al. (2002). Gene expression analysis using oligonucleotide arrays produced by maskless photolithography. *Genome Res* 12(11): 1749-55.
- Olayioye, M. A. (2001). Update on HER-2 as a target for cancer therapy: intracellular signaling pathways of ErbB2/HER-2 and family members. *Breast Cancer Res* 3(6): 385-9.
- Papin, J. A., T. Hunter, et al. (2005). Reconstruction of cellular signaling networks and analysis of their properties. *Nat Rev Mol Cell Biol* 6(2): 99-111.
- Perrimon, N., A. Friedman, et al. (2007). Drug-target identification in *Drosophila* cells: combining high-throughout RNAi and small-molecule screens. *Drug Discov Today* 12(1-2): 28-33.
- Phan, J., A. Zdanov, et al. (2002). Structural basis for the substrate specificity of tobacco etch virus protease. *J Biol Chem* 277(52): 50564-72.
- Ponten, J. and E. Saksela (1967). Two established in vitro cell lines from human mesenchymal tumours. *Int J Cancer* 2(5): 434-47.
- Romanov, S., A. Medvedev, et al. (2008). Homogeneous reporter system enables quantitative functional assessment of multiple transcription factors. *Nat Methods* 5(3): 253-60.
- Rossner, M. J. (2008). Combinatorial synthesis and use of libraries of short expressed nucleic acid sequences for the analysis of cellular events. EP 08 165 151.5. Prio.date: 25.09.2008.

- Sanger, F., S. Nicklen, et al. (1977). DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* 74(12): 5463-7.
- Schena, M., D. Shalon, et al. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270(5235): 467-70.
- Schlabach, M. R., J. Luo, et al. (2008). Cancer proliferation gene discovery through functional genomics. *Science* 319(5863): 620-4.
- Schoeberl, B., C. Eichler-Jonsson, et al. (2002). Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors. *Nat Biotechnol* 20(4): 370-5.
- Schulze, W. X., L. Deng, et al. (2005). Phosphotyrosine interactome of the ErbB-receptor kinase family. *Mol Syst Biol* 1: 2005 0008.
- Shendure, J. (2008). The beginning of the end for microarrays? *Nat Methods* 5(7): 585-7.
- Shendure, J. A., G. J. Porreca, et al. (2008). Overview of DNA sequencing strategies. *Curr Protoc Mol Biol* Chapter 7: Unit 7 1.
- Silva, J. M., K. Marran, et al. (2008). Profiling essential genes in human mammary cells by multiplex RNAi screening. *Science* 319(5863): 617-20.
- Stagljar, I., C. Korostensky, et al. (1998). A genetic system based on split-ubiquitin for the analysis of interactions between membrane proteins in vivo. *Proc Natl Acad Sci U S A* 95(9): 5187-92.
- Starkuviene, V., U. Liebel, et al. (2004). High-content screening microscopy identifies novel proteins with a putative role in secretory membrane traffic. *Genome Res* 14(10A): 1948-56.
- Suter, B., M. J. Fetchko, et al. (2007). Examining protein protein interactions using endogenously tagged yeast arrays: the cross-and-capture system. *Genome Res* 17(12): 1774-82.
- Suter, B., S. Kittanakom, et al. (2008). Two-hybrid technologies in proteomics research. *Curr Opin Biotechnol* 19(4): 316-23.
- Treisman, R. (1986). Identification of a protein-binding site that mediates transcriptional response of the c-fos gene to serum factors. *Cell* 46(4): 567-74.
- Wallasch, C., F. U. Weiss, et al. (1995). Heregulin-dependent regulation of HER2/neu oncogenic signaling by heterodimerization with HER3. *EMBO J* 14(17): 4267-75.
- Waterman, H., I. Alroy, et al. (1999). The C-terminus of the kinase-defective neuregulin receptor ErbB-3 confers mitogenic superiority and dictates endocytic routing. *EMBO J* 18(12): 3348-58.
- Waterman, H., I. Sabanai, et al. (1998). Alternative intracellular routing of ErbB receptors may determine signaling potency. *J Biol Chem* 273(22): 13819-27.
- Wehr, M. C., R. Laage, et al. (2006). Monitoring regulated protein-protein interactions using split TEV. *Nat Methods* 3(12): 985-93.
- Wehr, M. C., L. Reinecke, et al. (2008). Analysis of transient phosphorylation-dependent protein-protein interactions in living mammalian cells using split-TEV. *BMC Biotechnol* 8: 55.
- Zhao, Z. Y., Z. X. Yin, et al. (2007). Profiling of differentially expressed genes in hepatopancreas of white spot syndrome virus-resistant shrimp (*Litopenaeus vannamei*) by suppression subtractive hybridisation. *Fish Shellfish Immunol* 22(5): 520-34.