CHAPTER 17

Molecular tools in drug research – translational medicine

Sandra Eder and Volker Wacheck

Department of Clinical Pharmacology, Medical University of Vienna, Vienna, Austria

1 Introduction

The advent of high-throughput and increasingly sensitive research techniques in molecular biology in the last 20 years has altered fundamentally our understanding of molecular biology. The unraveling of the human genome has provided unprecedented insights into the molecular pathophysiology of diseases. DNA sequencing techniques allowed collecting huge amounts of genetic information, revealing genetic variation and variable expression of genes pinpointing to the molecular level of diseases. In parallel, analytical methods for evaluating large sets of proteins become available facilitating studies on the functional relevance of transient or stable expression of these molecules for a disease phenotype. These novel molecular techniques are nowadays at the heart of modern drug research as they allow identifying and validating novel molecular targets of disease, drug screening as well as the discovery of biomarkers for predicting and monitoring response to drug therapy.

To take advantage of this gain of information and to bring it into clinical practice "translational medicine" emerged as an "interface" research discipline. Translational medicine is deemed to link preclinical (= basic) science and clinical science. It is a research field, in which many science disciplines are integrated. Physical-, chemical-, biological-, biochemical- and immunology knowledge as well as a solid understanding of clinical medicine is required to get involved in the research about the molecular basis of pathologies and the development of concepts to alter the aberrant molecular signalling by therapeutic interventions. Following the "bench-to-bedside" concept, translational medicine starts with basic laboratory research. Physician-scientists work-

Keywords: Translational medicine, target identification, target validation, target deconvolution, SAGE (serial analysis of gene expression), cloning, sequencing, phage display, transfection, yeast two hybrid, mass spectrometry, knock-out animal, Cre/lox system, FISH, PCR, FACS, Blotting, Western/Northern, ACE2

ing at the interface between the research laboratory and patient care or a team of basic and clinical science investigators identify and explore the potential of novel molecular targets for a specific disease by the tools of molecular biology *in vitro*. After faithful validation of the biological relevance of the target, drug candidates are tested in cell culture based *in vitro* assays and *in vivo* animal models for their potential to modulate the identified pathological signalling pathway and their therapeutic potential. Given positive results, translational medicine transfers the basic laboratory discoveries into clinical "proof-of-concept" trials for improving patient-oriented treatment or prevention of a specific disease.

It is important to emphasize that translational medicine is not a "one-way" directed process rather than a perpetually stimulating circulatory effort. Besides "bench-to-bedside" the inverse "bedside-to-bench" research activities are a tremendous fruitful source of novel information of the underlying molecular pathophysiology of a disease. Fluids (e.g. blood, urine or cerebrospinal fluid) or tissue samples collected from patients are analysed by the tools of molecular biology in more and more detail for biomarker discovery. The availability of the novel – omics technologies in combination with biobanks, that is large collections of samples from patients with well-characterized clinical history, allow linking molecular expression patterns ("molecular fingerprint") with a disease state. Lessons learnt from these re-translational efforts provide the origin for novel research hypotheses to be tested then in preclinical. This illustrates that a close collaboration between clinical and pre clinical science, such as biochemistry, physics, chemistry, molecular biology and medicine is key for successful translational medicine.

The molecular tools available today in concert with translational medicine pave the way for unprecedented opportunities for drug research. These research activities will have great impact for understanding and combating diseases such as malignancy and infectious and autoimmune disease, as well as cardiovascular, metabolic and neurological diseases. In the following we will highlight exemplarily molecular tools and techniques employed in drug research for target identification, target validation, drug discovery and monitoring pharmacodynamic biomarkers in translational clinical trials.

2 Molecular tools for drug target identification

The identification of novel molecular targets in drug research may derive from different conceptual approaches. With the advances in molecular biology and biochemistry novel techniques emerged allowing whole-genome wide comparative analysis of patient and samples from healthy volunteers. Samples collected (e.g. tissue, serum and urine) are analysed for differential expression of genes and proteins. Differentially expressed genes (up- or downregulated) will then further characterized for their biological function and

their functional relevance. This "target screening" approach was clearly expedited by the availability of various molecular techniques in the last decade.

Alternatively, there are strategies starting from an observed phenotype induced by a drug candidate, which then will be further characterized to identify the target modulated. These so called "target deconvolution" strategies start form a given compound library. The compounds in such libraries are studied typically in mammalian cell culture high-throughput screening (HTS) assay systems for a phenotype of interest caused by the test compound. Depending on the drug effect desired, such phenotype might be morphological changes of the cell (e.g. cell shape, neurite outgrowth) or specific cellular effects, which will be measured by biochemical assays. For instance, cell viability measured by specific dyes relying on the metabolic ability of cells to reduce formazan salt, are employed to screen the cytotoxic activity of potential novel anti-cancer drugs. Reporter gene assays are employed to detect the activation of a particular signalling pathway of interest. For this purpose an easily detectable reporter gene (e.g. luciferase) is fused to the promoter sequence of a downstream target gene of the studied pathway. In case a compound screened interacts with the specific signalling pathway, the expression of the report gene will be activated or inhibited, which would become detectable by a change of luminescence. In the following, molecular tools employed for "target screening" and "target deconvolution" will be presented.

2.1 At the nucleotide level

As far as it concerns gene expression, the *DNA micro-array* technology represents a convenient, versatile and (in the meantime) affordable tool to assess the mRNA expression in samples. Gene expression profiles derived from mRNA micro-array may identify genomic signatures pinpointing to signalling pathway altered in a specific disease state and lead to the identification of new molecular targets. Moreover, micro-array may be used to link gene expression signatures from small molecules. By so-called "connectivity maps" gene expression profiles from human cell cultures treated with small molecules were linked successfully to identify molecules with a common mechanism of action. The micro-array technology is described into detail in Chapter 15. Briefly, by hybridization of the fluorescently-labelled "target" mRNA (sample) to the oligonucleotide probe sequences spotted on the array, the mRNA in a sample can be quantified by means of fluorescence intensity, which is proportional to the abundance of the target sequence in the sample. A limitation of the micro-array technology is the fact that it is by nature a biased approach. Since only these mRNA transcripts can be detected, which have complementary oligonucleotide probe sequences on the array, any further transcripts will not be detected. This might limit the use of micro-arrays for finding novel drug targets.

As an alternative, the *serial analysis of gene expression* (SAGE) provide an unbiased approach to measure the number of mRNA transcripts, since it is inde-

pendent on prior knowledge of what transcripts to study. SAGE is an effective method for analysing mRNA gene expression, which makes use of short cDNA fragments, so called sequence tags. These small sequences of nucleotides can effectively use to identify the original mRNA transcript. The gene tags are linked together, sequenced

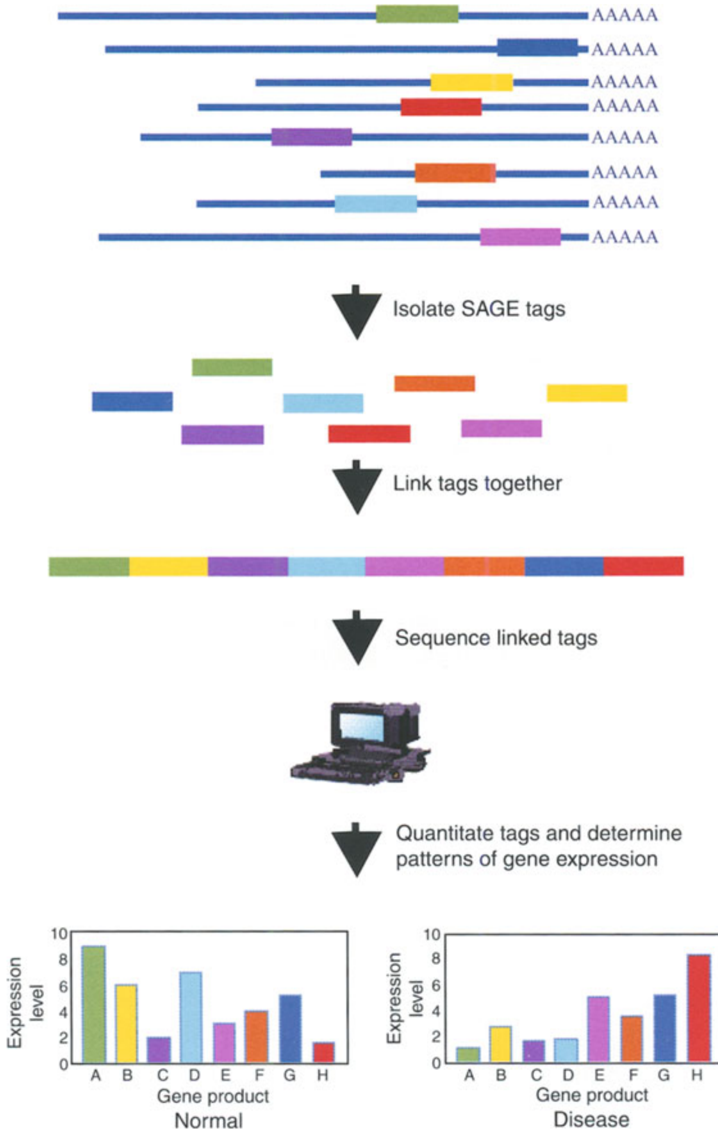

Fig. 1 Scheme of serial expression of gene expression (from http://www.sagenet.org/findings/index.html)

and numbered to deduce the quantitative expression of the original mRNA expression in a sample.

The first step of SAGE is extraction of the mRNA from a sample to be studied for its mRNA transcription levels (Fig. 1). The mRNA will be transcribed by the enzyme reverse transcriptase into cDNA (copyDNA), which is much more stable than mRNA. The total of the transcribed mRNA of a sample (e.g. tumour biopsy) is referred to as cDNA library as it comprises all transcribed sequences of a sample. Next, short sequence fragments ("tags") of 26–28 basepairs are cleaved from the cDNA strands by restriction endonuclease enzymes cutting the cDNA at defined nucleotide motifs. The cut cDNA fragments are amplified by PCR and subsequently linked to each other to form one long string of "tags" (called "concatemer"). This chain of tags is then introduced into a vector to be cloned.

Cloning is the process of producing multiple copies of a defined DNA sequence. It allows amplification of DNA fragments. Basically, the isolated DNA fragment to be amplified (= "insert"; in case of SAGE the concatemer) will be ligated enzymatically into a plasmid, which is an extra chromosomal DNA molecule capable to replicate independently from the chromosomal DNA (Fig. 2). Next, the plasmid with the inserted cDNA will be introduced into bacteria (for example *E. coli*) for propagation. Since this process of transformation yields typically rather low efficiency, plasmids contain a selection marker conferring antibiotic resistance. Only bacteria successfully transformed with the plasmid will grow on media containing this antibiotic. After forming single colonies, bacteria will be picked and further cultured before finally the amplified DNA fragment will be extracted from the bacteria.

To identify and quantify the "tags" in the cloned concatemer it will be finally sequenced. *Sequencing* is a method to determine the order of nucleotides bases in a molecule of DNA. The first techniques for gene sequencing were introduced in the 1970s by 2D-chromatography or the wandering-spot method. Later, the chain-termination method developed by Sanger (who received the Nobel Price for its development) became the method of choice. The Sanger method is based on the use of dideoxynucleotide triphosphates (ddNTPs) as DNA chain terminators. The DNA sample to be sequenced is analysed in four separate sequencing reactions. For each reaction all four of the standard deoxynucleotides (dATP, dGTP, dCTP and dTTP) are added but only one of the four dideoxynucleotides (ddATP, ddGTP, ddCTP or ddTTP), which are the chain-terminating nucleotides. By addition of DNA polymerase DNA strand extension starts until incorporation of a modified nucleotide terminates DNA strand elongation resulting in DNA fragments of varying length. By gel electrophoresis running each of the four reactions in a separate lane the DNA bands can be visualized and the relative position of the DNA bands translates into the DNA sequence. In extension of the Sanger method, dye-terminator sequencing employs four different fluorescent dyes, which permits sequencing in a single reaction instead of running four reactions in parallel. The fluorescently labelled ddNTPs and

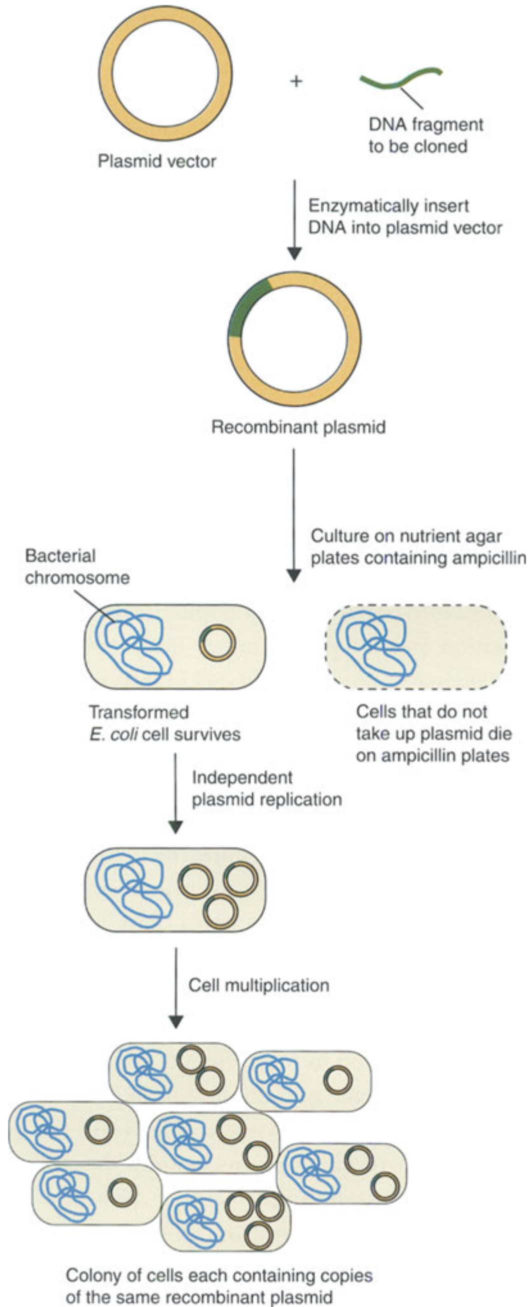

Fig. 2 Cloning of a DNA fragment into a plasmids (from http://www.ncbi.nlm.nih.gov/bookshelf/br.fcgi?book=mcb&part=A1582&rendertype=figure&id=A1590)

primers paved the way for automated, high-throughput DNA sequencing device employed today.

After sequencing, the tags within the concatemer can be identified by their sequence and attributed quantitatively to their original gene. The sequence of the tags will be correlated by means of a "sequence-similarity search" to the original mRNA and gene, respectively. Due to the fact that SAGE is a sequence-based technique, it is more accurate for quantifying gene expression compared to the hybridization based DNA micro-arrays. At the downside, SAGE experiments are much more work-intensive than micro-array studies and therefore the latter are preferred for large scale screening studies.

2.2 At the protein level

Proteomic is the analysis of proteins starting from the systematic separation and identification of all proteins within a cell, tissue, or other biological sample. For drug research, proteomic techniques are employed for "target screening" as well as for "target deconvolution".

In analogy to the DNA micro-arrays, *protein-arrays* have become available. Protein-arrays (also termed protein-chips) allow detecting antibody–antigen, protein–protein, protein–DNA or protein–small molecule interaction. There are different types of protein arrays. For "target screening" approaches protein lysates from a sample of interest (e.g. tumour biopsy) are putted on an "antibody micro-array" with a library of specific capture antibodies immobilized on its surface. Proteins in the probe will be captured by the spotted antibodies and subsequently labelled by a second fluorescence antibody for detection. Alternatively, the lysate probe itself can be spotted on the array and antibodies will be added. The binding reaction in these "reverse phase protein micro-arrays" is again detected by a fluorescently labelled second-step antibody. The beauty of the "reverse phase protein micro-arrays" is that they allow to test simultaneously many different lysates on a single chip. For target deconvolution, "antibody-arrays" performed from cells exposed to a drug candidate permit identification of protein signature profiles pinpointing to the potential mechanism of action in short time. Moreover, "reverse transfected cell micro-arrays" enable to evaluate in a high-throughput manner the interaction of proteins and drug candidates. Such micro-arrays are spotted with cDNAs from a library in expression vectors and overlaid with a transfection reagent and mammalian cells to generate a "living" micro-array. Cells are transfected with the spotted cDNA vectors and start expressing the respective proteins. Incubation with a labelled small molecule permits detection of cells interacting with the small molecule due to expression of the transfected protein. The beauty of the concept is that "Reverse transfected cell micro-arrays" allow screening proteins derived from any cDNA library without necessitating production of individual purified proteins.

A further method to identify proteins, their posttranslational modifications and interaction with other molecules is given by *mass-spectrometry*. Mass-spectrometry is an analytical technique based on measuring the mass-to-charge ratio of molecules. Biological samples often comprise a complex mixture of proteins requiring fractionation before analysis by mass spectroscopy. For this purpose, proteins are separated by two-dimensional gel electrophoresis (2DE) by high performance liquid chromatography (HPLC) after enzymatic digestion. Such fractionated peptides are then assessed by mass-spectrometry such as MALDI-TOF (matrix-assisted laser desorption–ionization–time of flight mass spectrometry). A laser beam passes through the sample to be analysed and causes vaporization and ionization of the sample. Ionized molecules fly upward into a tube. "Time of flight" through the tube correlates directly to mass: with lighter molecules having a shorter time of flight than heavier ones. Mass-spectrometry is used for qualitative and quantitative analysis of proteins.

In order to study physical drug–target interaction expression-cloning-based techniques are of particular interest. Recombinant proteins are expressed from cDNA libraries and exposed to the drug candidate to identify proteins interfering with the drug candidate.

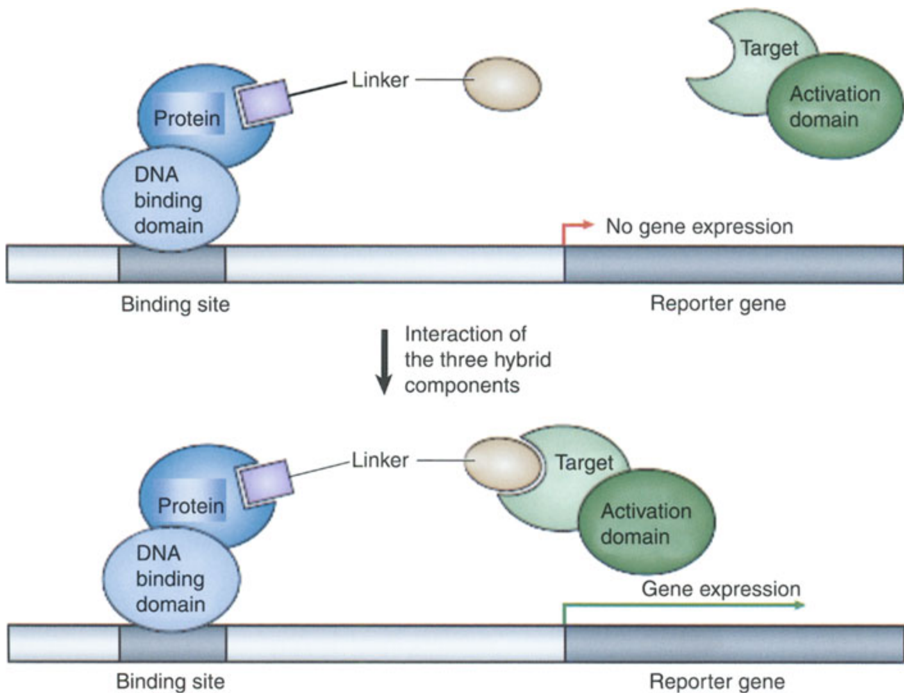

Fig. 3 Yeast three hybrid system (adapted from Ref. [1])

The *yeast two-hybrid* (Y2H) (see Fig. 3) is a molecular tool in biology to assess protein–protein or protein–DNA interactions. This technique was first designed to study protein–protein interaction in yeast (therefore "yeast" two-hybrid). In essence, it is based on the activation of a downstream reporter gene upon binding of a transcription factor to an upstream activating sequence. By splitting the transcription factor into a DNA binding domain and an activating domain, the interaction of a protein coupled to the activating domain ("prey") and a protein coupled to the DNA binding domain ("bait") brings the two fragments of the transcription factor in close proximity resulting in transcriptional activation of the reporter gene. For target deconvolution of drug candidates, the Y2H technique was adapted to a "yeast three-hybrid" technique by putting in between the two interacting proteins the drug candidate ("third hybrid"), which is coupled by a linker to the "bait" protein. Only if the target protein binds to the drug candidate reporter gene expression is activated.

Based on the reporter gene chosen a change in cellular phenotype is observed, allowing selection of cells expressing the protein interacting with the drug candidate. For generation of target "prey" proteins, random libraries or cDNA libraries (e.g. from a patient sample) are employed, which are ligated into plasmids to express the respective fusion protein in yeast. Y3H systems have successfully employed to deconvolute the targets of small molecule kinase inhibitors. However, Y2H (as well as Y3H) systems are prone to false positive results. Among other causes, this is due to overexpression of the fusion protein (= non-specific interactions) and the use of yeast as primitive organisms. To overcome, a mammalian cell based derivative was only recently introduced (MASPIT).

Phage display is another molecular biology technique to screen protein interactions drug discovery makes use of. Basically, cDNAs from a library are cloned into bacteriophages ("bacteria infecting viruses") to produce fusion proteins of the bacteriophage coat protein and the proteins corresponding to the respective cDNAs. This phage display library is then exposed to the target molecules immobilized on a surface. Only phages (short for bacteriophages) expressing proteins interacting with the target molecules will be eluted by washing steps and transfected in bacteria for amplification. The cDNA of the enriched phage population will be isolated and sequenced to identify the target protein. Phage display have been primarily employed for antibody lead selection but are also particularly well suited to identify new ligands (e.g. receptor blocker or agonists, antibodies) for low-abundance proteins and targets of low-affinity ligands.

3 Molecular tools for target validation

Following target identification by "target screening" and "target deconvolution" it is essential for drug development of molecular targeting drugs to confirm that the target

identified is of functional relevance for the phenotype of the disease of interest. Despite thoughtful target identification strategies applied, it might turn out that the target identified is a false positive one, in that the target indeed is expressed in a given disease state but its modulation does not result in any change of phenotype ("epiphenomenon").

To accomplish this task there are several molecular tools and model systems available. In principle, the identified molecular target will be either (over)expressed to promote the phenotype or down-regulated to suppress the phenotype of a disease with the aim to foster the causal relationship between the molecular target and the disease.

3.1 Target (over)expression

The idea of *transfection* studies is to transform a well characterized mammalian cell line to express the gene of interest. The phenotype of the generated new cell clone will be evaluated in functional assays in comparison to the original cell line and any change in phenotype will be attributed to expression the gene of interest. For successful transfection experiments it is essential to bring as much genetic copies of the recombinant DNA in as many cells as possible in a way that preferably many cells "survive" the transfection procedures.

To introduce a transgene into cells typically plasmids will be employed as vectors. As described above, plasmids are extra-chromosomal DNA molecules, which are separated from the chromosomal DNA and can therefore replicate autonomously. The gene of interest will be inserted into the plasmid at the so called "Multiple Cloning Site", which consists of restriction sites allowing enzymatic cleavage of the plasmid, insertion of the transgene and re-ligation of the plasmid with the inserted cDNA encoding for the gene of interest. For replication, vectors have an "origin of replication" (= sequence of DNA capable of directing the propagation of itself and any linked sequence) for initiating semi-independent replication of the plasmid in host cells.

Mammalian cells are protected by their cell membrane from potential environmental risks such as foreign DNA. In order to surmount the lipid bilayer and facilitating the uptake of DNA into the host cell a number of different transfection methods are available in molecular biology. The "classical" approach of transfection is based on a transient increase in the permeability of the cell membrane. Calcium phosphate forms precipitates with DNA, which are then taken-up by cells. Cationic lipids and polymers bind the negatively charged DNA and cellular uptake is mediated by fusion of the complex with the cell membrane or endocytosis, respectively.

The most efficient and reliable approach for transfection of mammalian cells is by the use viral vectors. Formally, the transfer of DNA into mammalian cells by use of viral vectors is referred to as "transduction" in contrast to the chemical based methods mentioned above. Viral mediated transduction strategies consist of a recombinant viral

vector/plasmid with the inserted gene of interest, which is transduced into a complementing packaging cell line to produce infectious virions (= extracellular virus particle). The virions are harvested and employed to transduce the target cells. The resembling of the naturally way of virus infection leads to high yields with near to 100% of transduced target cells. Whereas adenovirus based system result in transient transduction of cells, retro- and lentiviral systems are integrated into the DNA of the target cells leading to stable expression of the insert in dividing mammalian target cells. For transduction of non-dividing and other cells not easily to transfect, lentiviral systems are considered to be the most promising vectors for. To avoid any uncontrolled virus spread viral transduction systems typically employ only replication incompetent viruses lacking essential genes necessary for generation of structural proteins. Still, the conduct of viral transfection experiments requires biosafety level II laboratory facilities. Given the considerable advantages of viral transfection systems they have become the standard for target validation studies *in vitro*.

3.2 Target downregulation

A point of criticism inherent to any target overexpression strategy for validation experiments arise about the question whether a target gene overexpressed in a mammalian cell line properly reflects the biological function of the endogenously expressed gene. Overexpression of a transgene might overwrite the fine-tuned balance of intracellular signalling leading to a phenotype aberrant from the one caused by expression of the endogenous target gene. Not only, but also for this reason the reverse strategy of target downregulation is often employed for target validation. By targeted suppression of the gene of interest the resulting phenotype can be attributed to the gene's biological function.

For *in vitro* experiments, the most versatile tool for target silencing is given by *RNA interference*. RNA interference is a naturally existing cellular mechanism of post transcriptional gene silencing. Mello and Fire published in 1998 that gene specific silencing may be achieved by double-stranded RNA complementary to the sequence of the target gene leading to degradation of the target mRNA by induction of an endogenous RNA induced silencing complex (RISC; for details see Chapter 20). Degradation of the target mRNA deprives the cell of the template for its protein translation machinery. Thus, RNA interference allows gene specific silencing of virtually any gene of interest. For target validation *in vitro*, the RNA interference mechanism may be induced by two different technical approaches.

siRNA (= small or short interfering RNA) are short double-stranded RNA fragments, which are introduced intracellular by transfection. Transfection of a cell with siRNA leads to activation of the RISC and subsequent target-mRNA degradation. siRNA experiments are easy to perform straight forward experiments for target validation, but they result only in transient downregulation of target mRNA.

For long-term target silencing RNA interference may be induced by introducing a vector into cells encoding a "*small hairpin RNA*" (sh-RNA). sh-RNA is a RNA sequence of about 70 nucleotides consisting of a sense – and a complementary antisense strand, which are connected by a short sequence of nucleotides. Since the sense- and the antisense strand tend to form a double strand by complementary base pairing with a loop at the top formed by the nucleotide sequence connecting the both strands the construct resembles the shape of a hairpin. The transcribed sh-RNA is exported from the nucleus to the cytoplasm where it is sliced by an enzyme called dicer into siRNA fragments, which then in turn activate the RISC for target mRNA degradation. RNA interference by siRNA is currently mainly limited to *in vitro* experiments since the delivery of siRNA molecules *in vivo* has not yet been adequately addressed to obtain reliable target silencing *in vivo* (despite in the liver).

An alternative strategy for studying the biological relevance of a target gene is the use of *knock-out* animals. Knock-out animals (typically mice) are genetically modified

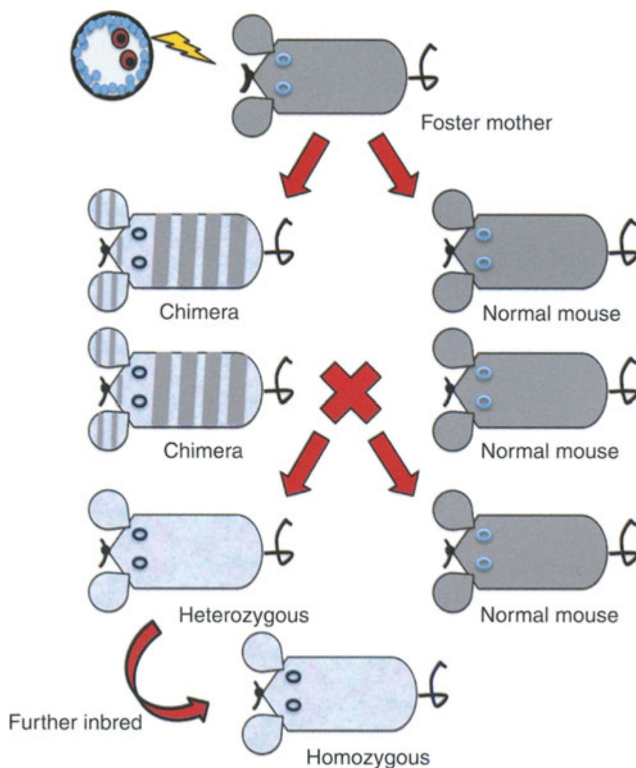

Fig. 4 Generation of knock-out mice

animals (GMA) in which one or multiple genes are deactivated by genetic manipulation of the animal's germ line. Embryonic stem cells are isolated from a mouse-blastocyst and propagated *in vitro*. By microinjection with a tiny glass pipette a DNA sequence is introduced into the nucleus of an embryonic stem cell. The introduced DNA sequence is highly similar (homologous) to the target gene sequence, except that by mutations the sequence is coding for a non-functional copy of the target gene. In some embryonic stem cells this nucleotide sequence is incorporated into the DNA by replacing the original copy of the gene (= homologous recombination). Following selection the successfully transformed embryonic stem cells are inserted into a mouse blastocyst and implanted into a uterus of a female mouse (see Fig. 4). The offspring of this foster mouse have a "mixed" genotype (= chimera) since they carry both, the original "wild-type" gene from the stem cell of the foster mouse and the "knock-out" gene from the engineered stem cells. By back-crossing these chimeric mice with wild-type mice heterozygous animals are obtained (= two different alleles of the gene of interest). Further inbred of these heterozygous mice results in homozygous knock-out mice with no copy of the functional gene of interest.

Since the first knock-out mouse models created in 1989, there are in the meantime several thousand strains of knock-out mice available for studies of diseases such as cancer, metabolic- or neurologic disorders (see http://www.knockoutmouse.org). Knock-out mice have become a valuable and reliable tool for validating the functional relevance of a gene of interest. Even simultaneous knock-out of two (= double knock-out) or three (= triple knock-out) genes and allele-specific knock-out to create homo- and heterozygous mice is feasible for studying the interaction and functional dependence of genes.

In comparison to gene silencing by sh-RNA, knock-out animals are in general more appropriate to study the phenotype of a complete suppression of a target gene ("knock-out"). In contrast, animals with gradual expression of sh-RNA constructs allow studying quantitative target-effect relationship for drawing conclusions about the degree of inhibition necessary for influencing the disease phenotype. Despite the attractiveness of knock-out models, it needs to be emphasized that like overexpression also gene silencing/knock-out provides only an approximation to the human situation. There are examples in the literature where the phenotype of a knock-out mouse displayed different characteristics relative to the loss of function of the same gene in humans (e.g. p53 knock-out mice), underlining the genetic inter-species differences between mouse and men. A further limitation of knock-out models stems from the fact that a substantial number of genes are of critical importance during embryonic development. In about 15% of all cases gene knock-out causes embryonic lethality of the transformed embryo. Since genes may serve different functions during embryogenesis and in adult animals, such gene might still be relevant for the phenotype of a human disease in adult mice but knock-out models are under such circumstances no way to validate this gene as drug target.

3.3 Conditional target regulation

In order to circumvent the limitation of constitutive knock-out animals with permanent gene knock-out animal model with conditional target regulation have been developed. Conditional target regulation facilitates expression or knock-down of a target gene limited to restricted tissues or only during a specific timeframe by employing tissue specific promoter and inducible systems, respectively. According to the course of a disease, a gene of interest can be switched off and on any time allowing dynamic studies for the biological relevance of the target over the course of disease.

In this context, the *Cre/lox system* is widely used. The Cre–lox system is a site specific recombination system for enzyme mediated cleavage and ligation at defined DNA sequences. It consists of the Cre (which stands for "Cyclization *Re*-combination") recombinase protein and the loxP ("locus of X-over P1") recombination sites containing a binding site for Cre. To knock-out a gene of interest, loxP sites are inserted in the DNA sequence before and after the sequence encoding the gene of interest (also referred to as "to flox a gene"). The Cre recombinase binds to the inserted loxP sites, cuts out the double strand between the two loxP sites and recombines the DNA ends at the loxP sites. This will result in a deletion of the gene of interest flanked by the loxP sites. For animal models, Cre protein and loxP sites are introduced separately as transgenes in mice. The offspring of these mice express then a complete functional Cre/lox system.

For conditional target regulation the expression of the Cre protein can be set under control of a tissue specific promoter. A tissue specific promoter expresses the genes under its control only in defined tissues. For instance, Cre protein under control of liver specific promoter allows the knock-down of a floxed target gene only in the liver of an animal. Likewise, tissue specific promoters are also employed to express a gene in a tissue restricted fashion to evaluate the "overexpression" phenotype of a gene of interest.

For time restricted expression of a gene of interest the Cre/lox systems has been adapted. The Cre protein was fused with a modified ligand binding domains of the estrogen receptor, which is no more activated by endogenous estrogen but by the selective estrogen modulator tamoxifen. By feeding animals with tamoxifen, tamoxifen binds to the modified estrogen receptor and induces Cre protein expression leading to deletion of the "floxed" target gene. Upon stop of tamoxifen administration the knock-out phenotype can be reverted.

Another approach for conditional target expression is given by the *tetracycline controlled transcriptional* (Tet) activation system. Expression of the gene of interest is controlled by tetracyline dependent promoters, which upon presence of externally administered tetracycline (or doxycycline as the more stable tetracycline analogue) either induce target gene expression (Tet-On) or indirectly repress target gene expression (Tet-Off). Relative to (inducible) Cre/lox systems, the Tet system enables

tighter control of target gene expression but requires more time before change in target expression due to induction/set-off becomes effective.

4 Molecular tools for monitoring pharmacodynamics in translational studies

After successful target identification/deconvolution and validation, a drug candidate undergoes GMP production and toxicology testing before it is allowed to be studied finally in first in men clinical trials. Given a molecule with a defined mechanism of action it is of outstanding importance in these early clinical trials to learn, beyond clinical tolerability, whether the postulated mechanism of action "translates" into the human setting ("bench-to-bedside"). In such "proof-of-concept" studies the changes induced by the drug candidate on the molecular level of the disease needs to be closely monitored by pharmacodynamic analyses and correlated with the pharmacokinetics of the drug candidate (PK/PD studies). Pharmacodynamic "biomarker" will be measured in samples from the tissue involved in the pathogenic process or from a surrogate matrix (e.g. normal skin, blood, urine) before and after treatment. Target modulation as an effect of the drug candidate administered will be assessed by molecular analysis. For monitoring pharmacodynamic biomarkers, a wide range of molecular tools for mRNA-, protein-analyses, and functional assays is available. Depending on the molecular target addressed and the accessibility of the target tissue either the direct – (e.g. down-regulation or phosphorylation of the target) or an indirect effect on the molecular target ("downstream" effect) will be measured.

The molecular tools applied for monitoring pharmacodynamics in translational studies are greatly overlapping with the ones employed for target identification and validation. However, in contrast to the – omic approaches used for target identification/ deconvolution the techniques applied for biomarker assessment in clinical trials are more focused (i.e. target-orientated), not at least due to the demanding standards required for biomarker assay validation.

At the nucleotide level, *fluorescence in-situ-hybridization* (FISH) is a standard molecular tool to detect DNA or mRNA in cells or tissue samples. FISH is based on binding of a labelled DNA oligonucleotide ("probe") complementary to the DNA or mRNA sequence of interest. Following denaturation of the DNA in the sample by heating the "probe" binds to the target DNA and is detected by its fluorescent label with fluorescence microscopy.

While FISH allows qualitative analysis, quantitative analysis of mRNA expression may be accomplished by *Northern blotting*. The technique of Northern blotting was first described in 1977 as a derivative of southern blotting used for DNA detection. Following extraction of mRNA from a sample, RNA's are separated by size via gel electrophoresis. Next, RNA samples are transferred to a nylon membrane ("blotting") and fixed on the

surface of the membrane by heat or UV light. For detection of specific mRNA, a labelled probe (radiolabelled or chemoluminescence) with an oligonucleotide sequence complementary to the mRNA of interest will be added for hybridization. The binding of the labelled probe is visualized on a X-ray film and can be quantified by densitometry. Northern blotting is characterized by a high specificity but a low sensitivity and a somewhat poor performance for quantitative analysis of mRNA.

In contrast, the *polymerase chain reaction* (PCR) is a highly sensitive molecular method for detection of mRNA in a patient sample. PCR multiplies the mRNA of interest by repeated amplification cycles allowing also detection of low abundant mRNA. Basically, mRNA of a biological sample is first reversely transcribed in more stable cDNA. Short sequences of DNA (so called "primers") hybridize to defined sequences of the cDNA of interest and initiate with an added polymerase selective extension and amplification of the cDNA of interest in cylces of repeated heating and cooling ("chain reaction"). For quantification of cDNA copies generated by PCR, real-time PCR (RT-PCR) intercalates fluorescent dyes in the PCR products so that the intensity of fluorescence detected directly correlates with the relative number of cDNA copies generated. RT-PCR (also coined q(uantitative)PCR) has become the most efficient tool for mRNA quantification in translational research studies.

While the transcriptome is by nature always only the "intention of protein expression", measuring target modulation at the protein level is considered to provide more reliable data in "proof-of-concept" studies. For assessment of protein based biomarkers there are several molecular methods available.

To analyse target protein expression in patient tissue samples, immunohistochemistry (IHC) is a well-established technique. Tissue slices are exposed to antibodies specific for the target of interest. Binding of the primary antibody to the antigen of interest is detected by a labelled secondary antibody, which reacts with the primary antibody. The secondary antibody is coupled with an enzyme catalyzing the reaction with a chromogenic substrate for visualization. The convenience of IHC for translational studies is given by the fact that it allows protein biomarker assessment of cells within their histological context. For multiplex histological analysis, "tissue micro-array" (TMA) have been introduced only recently, allowing to assess simultaneously up to 1000 small tissue sample cores embedded in a single paraffin block instead of staining all samples on separate slides. At the downside, quantification of IHC is observer dependent, requires skilled pathologists and is complex to standardize. Typically, results are presented as a mixed score consisting of the percentage of cells positive and gradual staining intensity for the biomarker in a limited number of regions of the sample. To overcome the observer and selection bias, digital tissue analysis system have been developed enabling automated acquisition and subsequent quantification of IHC stained sections ("tissue FACS").

For quantitative analysis of protein expression, *Western blotting* has been the molecular standard tool for many years. In contrast to Northern blotting, proteins

(not mRNA) are separated for size and charge by gel electrophoresis. Following transfer to a membrane ("blotting") the protein of interest are detected by specific antibodies. Binding of the specific antibody ("primary antibody") is visualized by a second step antibody binding to the primary antibody and catalyzing a "reporter reaction" (e.g. chemiluminescence or colorimetric reaction). This reporter reaction is detected as bands on an X-ray film whereby the intensity of the band correlates with the amount of protein in the sample. Western blotting as a standard laboratory tool is characterized by a satisfying specificity but it is considered as a semi-quantitative technique only and therefore only rarely employed as a biomarker assay for drug development.

Fluorescence-activated cell sorting (FACS) or flow cytometry enables quantitative detection of multiple proteins of individual cells in a standardized manner. In principle, while cells pass through a laser beam in a stream of fluid, the cells and their capacity to bind fluorescently labelled antibodies specific for the proteins of interest is quantified by detecting the scattering of light and the fluorescence excited by the light of the laser beam. Flow-cytometry is suited for all type of samples where cellular protein expression in cell suspensions is of interest (e.g. peripheral blood mononuclear cells). On a cautionary note, the specificity of FACS is purely dependent on the specificity of the antibodies employed, which needs to be validated carefully before a FACS assay may be employed as biomarker for decision making in translational studies.

To measure non-cellular protein expression in body fluids *Enzyme-linked Immunosorbent Assay* (ELISA) is employed. ELISA rest upon detection of antigens present in fluids such as serum, urine or cerebrospinal fluid by binding to specific antibodies. In a standard "sandwich" ELISA, specific antibodies immobilized on a surface of a microtiter plate capture the protein of interest in the sample fluid. After washing the plate for removing unbound proteins, a second antibody specific for the protein of interest is added, which is linked to an enzyme reacting with a subsequently added substrate for detection (e.g. horseradish peroxidase or alkaline phosphatase). ELISA might also reversely be designed to detect antibodies fluids by exposing the corresponding antigen in the assay or to detect secretory products of activated immune cells cultured in microtiter plates (i.e. ELISPOT). The most recent progress in ELISA development allows simultaneous detection of multiple antigen within one sample (up to 100 antigens) in so called Multiplex ELISA (= Luminex). These assays make use of colour coded beads ("microspheres") for labelling the individual reactions taking place simultaneously in the same well.

In addition to this general set of molecular tools outlined here, a multitude of techniques fitting the needs of particular research fields are currently emerging, which have the potential to impact substantially future drug research (for instance "circulating tumour cells" as "liquid biopsy" in oncology [2] or "circulating endothelial progenitor cells" for angiogenesis research [3]).

Case Study: Re-translational studies – the case of ACE2

As mentioned in the introduction, translational research may not be seen exclusively as a "one-way process" bringing basic research success stories into clinical trials but also as the inverse process by inspiring basic research with novel clinical relevant information for hypothesis forming and testing. A recent example illustrating this fruitful circle of "bedside-to-bench-to-bedside" in translational research is the development of the Angiotensin converting enzyme 2 (ACE2) recombinant protein currently in clinical testing.

The angiotensin converting enzyme (ACE) is part of the Renin-Angiotensin-Aldosterone-System (RAAS). It converts angiotensin I to angiotensin II and plays an important role in regulation of the cardiovascular system including blood pressure, blood volume and serum electrolytes. After its first description in 1956 [4] and the deciphering of its functional role in the 70s of the last century [5], ACE inhibitors blocking the enzymatic activity of ACE are in daily clinical use since 1981.

In 2000 it was published that there is a second ACE gene (ACE2) [6]. This human homologue was identified as a result of a re-translational research effort. From the explant of a female patient undergoing heart transplantation due to dilated cardiomyopathy tissue samples were collected. mRNA was extracted from the explanted cardiac left ventricle and reversely transcribed in cDNA. From this cDNA, a human cardiac left ventricle cDNA library was prepared by using standard cloning techniques. A total of 19,000 different clones were generated and sequenced by high throughput methods. The nucleotide sequences obtained were checked for homology with any already known gene sequence by means of a "sequence-similarity search". Among the 19,000 clones, a so far unknown gene sequence was identified showing 42% identity with the amino domain of the known ACE gene. In order to study the tissue distribution of this novel ACE2 gene, 23 different human tissue types were evaluated by northern blotting with ACE2 specific probes. ACE2 mRNA tissue expression differed from ACE expression with preferential expression in heart, kidney, testis and at a lower level in the colon and the lung. To define cellular and subcellular localization of ACE2 expression, immunohistochemical examination of human ventricular myocardium with an ACE2 specific antibody was performed. ACE2 protein expression was found to be localized to the endothelium of most intramyocardial vessels including capillaries and venules. In order to study the functional role of ACE2 expression, mammalian cells were transfected transiently with a vector encoding the human ACE2 gene. There was no apparent difference in phenotype of transfected and non-

transfected control cells. However, mass-spectrometry revealed that only in the supernatant of cells transfected with ACE2 Angiotensin I was converted into Angiotensin 1–9 and Angiotensin II into Angiotensin 1–7. This effect could not be inhibited by addition of the ACE inhibitor lisinopril. Thus, ACE2 differed in its enzymatic activity from ACE.

However, at this point in time the biological role of ACE2 was still unknown. To gain insight into the biological relevance of ACE2 a knock-out mouse was created [7]. Disruption of the murine ACE2 gene resulted in increased levels of Angiotensin II and progressive worsening of cardiac contractility with age. It turned out that ACE2 apparently counterbalance the function of ACE within the local RAAS. While ACE increases Angiotensin II level promoting diseases such as cardiomyopathy, ACE2 degrades Angiotensin II to Angiotensin 1–7 protecting from cardiomyopathy. The counterbalancing effect of ACE2 is also observed in other organs like the kidney. Deletion of ACE2 in Akita mice, a mouse model with a mutation causing a diabetic phenotype, leads to spontaneous onset of nephrotic glomerulonephritis and diabetic kidney injury [8].

A crucial role of ACE2 in the respiratory system emerged by the discovery that the SARS virus is able to bind to ACE2. In an elegant study it was shown that ACE2 is essential for SARS infections *in vivo*. ACE2 knock-out mice infected with SARS were protected from SARS virus replication in the lung [9]. At the same time, SARS virus replicating in lungs of normal wild-type mice leads to down-regulation of ACE2 protein expression in the lungs.

The SARS virus is one of many potential inducers of the most serious form of acute lung injury, the acute respiratory distress syndrome (ARDS). ACE2 knock-out mice suffering from acute lung injury induced by different stimuli show more severe symptoms in form of enhanced vascular permeability, increased lung edema, neutrophil accumulation and worsened lung function [10]. Of note, ACE2 is able to counteract these effects by playing a protective role in acute lung injury with therapeutic potential. In mice with acute lung injury, treatment with recombinant ACE2 protein improved the symptoms of acute lung injury [10]. This was observed in wild-type as well as and in ACE2 knock-out mice.

These multiple lines of evidence foster the validity of ACE2 as a promising molecular target for treating acute lung injury by recombinant ACE2 and provided the rationale for translating the concept into clinical testing. Only recently, a Phase I study in healthy volunteers was announced to be completed for clinical developing of recombinant human ACE2 as an enzyme biotherapeutic in patients with ARDS (http://www.apeiron-biologics.com).

References

1. Terstappen GC, Schlupen C, Raggiaschi R, Gaviraghi G (2007) Target deconvolution strategies in drug discovery. Nat Rev Drug Discov 6: 891–903
2. Pantel K, Alix-Panabieres C, Riethdorf S (2009) Cancer micrometastases. Nat Rev Clin Oncol 6: 339–351
3. Bertolini F, Mancuso P, Shaked Y, Kerbel RS (2007) Molecular and cellular biomarkers for angiogenesis in clinical oncology. Drug Discov Today 12: 806–812
4. Skeggs LT Jr, Kahn JR, Shumway NP (1956) The preparation and function of the hypertensin-converting enzyme. J Exp Med 103: 295–299
5. Ferreira SH, Greene LH, Alabaster VA, Bakhle YS, Vane JR (1970) Activity of various fractions of bradykinin potentiating factor against angiotensin I converting enzyme. Nature 225: 379–380
6. Donoghue M, Hsieh F, Baronas E, et al. (2000) A novel angiotensin-converting enzyme-related carboxypeptidase (ACE2) converts angiotensin I to angiotensin 1–9. Circ Res 87: E1–E9
7. Crackower MA, Sarao R, Oudit GY, et al. (2002) Angiotensin-converting enzyme 2 is an essential regulator of heart function. Nature 417: 822–828
8. Wong DW, Oudit GY, Reich H, et al. (2007) Loss of angiotensin-converting enzyme-2 (ACE2) accelerates diabetic kidney injury. Am J Pathol 171: 438–451
9. Kuba K, Imai Y, Rao S, et al. (2005) A crucial role of angiotensin converting enzyme 2 (ACE2) in SARS coronavirus-induced lung injury. Nat Med 11: 875–879
10. Imai Y, Kuba K, Rao S, et al. (2005) Angiotensin-converting enzyme 2 protects from severe acute lung failure. Nature 436: 112–116