1. ENTREZ search and PubMed search for relationship with key target or key phenotypic link

2. GENELIST I: Non-redundant gene list:
   - Ranking in descending order based on # references associated with each gene

3. GENELIST II: Non-redundant gene list:
   - Extraction of all individual genes cross-referenced in abstracts for each gene from GENELIST I
   - Ranking in descending order based on # of individual genes from GENELIST I referenced for each gene

4. SUBLIBRARY A: the top of ranked GENELIST II, Manual intervention/control of a fraction of genes

5. SUBLIBRARY B: All genes from GENELIST I (and II) with <3 references AND in top impact factor journals

6. SUBLIBRARY C: All genes from GENELIST I with <3 references and publ. date < 1 year => manual curation.

7. FINAL TARGET LIBRARY: A+B+C.

(57) Abstract: The present invention provides methods of using RNAi in a high throughput screening for the identification and validation of drug targets in a physiological, disease-relevant context.
— of inventorship (Rule 4.17(iv)) for US only

Published:
— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.
RNAI-BASED TARGET IDENTIFICATION AND VALIDATION

BACKGROUND OF THE INVENTION

1. Field of the Invention

[0001] The present invention relates to methods for identifying and validating targets of biologically relevant pathways. More specifically, the present invention relates to the use of RNAi to identify and validate drug targets in vitro and in vivo mammalian disease models.

2. Description of Related Art

   a. Drug Design

[0002] The average time to bring a drug to market has increased over the last 30 years from an average of 8.5 years to more than 13 years. At the same time, the average cost to bring a drug to market has increased to over $890 million as of 2003. More than 60% of this alarming rate of attrition is accounted for by lack of efficacy in phase II and/or phase III trials and toxicity, including mechanism-based toxicity. The average cost of a drug failing in phase II exceeds $50 million. In order to reduce the likelihood of more failures, there remains a need for a rational method to identify appropriate specific drug targets that are causally involved in a specific disease, and the consequences of modulating these causal, disease-associated molecular targets.

   b. Identification of Gene Function

[0003] Two major strategies are used to define gene function in a given organism: “forward genetics” relies on the identification of genes that cause a well-defined phenotypic change when mutated (from phenotype to genotype), while “reverse genetics” is based on the intentional alteration of expression or function of a known DNA sequence and observation of the resulting phenotype (from genotype to phenotype). Forward genetic screens are well suited for model organisms such as yeast, Drosophila and C. elegans, because their genomes can be easily and efficiently altered, allowing the identification of rare individuals with a desired phenotype from a large number of mutants. Indeed, random mutagenesis conducted in these organisms has led to the discovery of genes involved in fundamental biological processes, like cell cycle and division, programmed cell death and embryonic patterning.

[0004] In more complex experimental organisms, such as the mouse, reverse genetic strategies have been traditionally preferred to define gene function, because it is more difficult to
manipulate the genome of these species and to obtain large numbers of mutants. Gene ablation by homologous recombination is a potent reverse genetic tool and can be performed relatively easily in some experimental organisms, such as yeast, on a large-scale basis. However, the process is more laborious and time-consuming in mice, and allows the “knock out” of only a single gene at a time. As the complete genome sequences of many key experimental organisms, including that of the mouse, have become available, the demand has increased for high throughput reverse genetics technologies that allow the functional annotation of hundreds or thousands of genes.

c. RNAi


1) Primitive Organisms

[0006] Two landmark scientific breakthroughs in 1998 provided the basis for large-scale functional genomic screens: the C. elegans genome had been sequenced and RNAi was discovered in the same organism (Fire et al., Nature 391:806-811, 1998). Soon thereafter, the initial microinjection protocol developed to induce RNAi became quickly substituted with a more robust and efficient procedure: the worms were fed with bacteria over-expressing the desired dsRNA (Timmons et al., Gene 2001, 263:103-112; Timmons et al., Nature 1998, 395:854). This allowed the generation of dsRNA libraries and RNAi-based high throughput genetic screens (Kamath et al., Methods 30:313-321, 2003). The first large-scale screens focused on easily detectable phenotypic changes, like viability and sterility, and identified the biological role of a few hundred of genes located on chromosomes I and III (Gonczy et al., Nature 2000, 408:331-336; Fraser et al., Nature 2000, 408:325-330). More recently, a genome-wide RNAi screen targeting 86% of the predicted transcripts led to the functional annotation of an additional 1,700 genes (Kamath et al., Nature 2003, 421:231-237). These and other RNAi-based screens identified the role of genes that have conserved orthologs in mice and humans and are involved in processes as diverse as fat metabolism, mitochondrial function or embryogenesis (Lee et al., Nat Genet 2003, 33:40-48; Piano et al., Curr Biol 2000, 10:1619-1622; Piano et al., Curr Biol 2002, 12:1959-1964; Ashrafi et al., Nature 2003, 421:268-272.)
[0007] Long, double stranded RNA molecules can also efficiently silence gene expression in another popular model organism, the fruit fly (Kennerdell et al., Cell 1998, 95:1017-1026; Misquitta et al., Proc Natl Acad Sci USA 1999, 96:1451-1456). Drosophila is particularly relevant for human biology, since more than 60% of disease-related genes have homologs in this animal. Moreover, established Drosophila cell lines can easily take up dsRNAs from the culture medium (Caplen et al., Gene 2000, 252:95-105; Clemens et al., Proc Natl Acad Sci USA 2000, 97:6499-6503). This observation and completion of the Drosophila genome project a few years ago allowed the initiation of systematic loss-of-function screens in insect cell lines that focused mostly on important signal transduction cascades, like the Wnt-, Hh-, PI3K- and MAPK-pathways (Clemens et al., 2000; Muda et al., Biochem J 2002, 366:73-77). A dsRNA-library targeting 43% of the Drosophila genes including all known kinases and phosphatases has recently been synthesized (Lum et al., Science 2003, 299:2039-2045). In another high throughput screen using RNAi, approximately 1000 predicted cell shape-regulatory molecules were targeted to detect well-defined morphological phenotypes, which correlated with the silencing of 160 genes (Kiger et al., J Biol 2003, 2:27).

[0008] In summary, genome-wide RNAi screens in C. elegans and Drosophila have become feasible approaches for the functional annotation of hundreds of genes. Where these organisms express homologs of human genes, this approach has provided important insight into the possible mode of action of disease-associated genes. However, the use of C. elegans and Drosophila as testing grounds for normal and diseased function of human tissues is limited by the greater degree of complexity of vertebrate genomes and physiology. This has stimulated the development of RNAi-based strategies to assay gene function directly in mammalian cells and tissues.

2) RNAi screens in mammalian tissue culture cells

[0009] Initial application of RNAi to mammalian cell culture systems was hampered by the fact that long dsRNA molecules induce an antiviral response that results in a global (non-specific) shutdown of protein synthesis. The discovery that approximately 21-23 bp long, so called short interfering RNAs (siRNAs) could induce sequence-specific gene silencing without activating the interferon pathway allowed the use of RNAi for functional genomics in mammals (Elbashir et al., Nature 2001, 411:494-498; Caplen et al., Proc Natl Acad Sci USA 2001, 98:9742-9747).
[0010] High throughput RNAi-based screens have been performed in human cell culture systems allowing the identification of novel components in important pathways involved in the control of cell death. De-ubiquitinating enzymes have been targeted by using plasmid-mediated shRNA delivery and detecting induction of NF-κB signaling, as measured by a luciferase reporter (Brummelkamp et al., Nature 2003, 424:797-801), which identified the biological role of a tumor suppressor gene involved in familial cylindromatosis. In another study, 510 genes were targeted, including 380 kinases, to identify their role in TRAIL-induced apoptosis (Aza-Blanc et al., Mol Cell 2003, 12:627-637), which identified genes with previously unknown function that were accelerating or inhibiting TRAIL-induced apoptosis. A lentiviral RNAi library covering close to 10,000 human and more than 5,000 mouse genes has been reported (Paddison et al., 2004).

[0011] These examples demonstrate that identification of therapeutically important genes by RNAi-based phenotypic screens can be achieved in mammalian cells. However, cell culture systems are not applicable for modeling complex physiological conditions. As a result, drug targets identified by current RNAi screening methods are prone to the same attrition rate in drug development as previously mentioned. Therefore, a continued need exists for the development of improved methods for identifying and validating drug targets for mammalian, in particular human, diseases.

3) Knock Down Mice using RNAi

[0012] While high throughput forward genetic screens has been done extensively in lower organisms, such as yeast, C. elegans, and Drosophila, this type of screens has hitherto not been feasible in mammals, such as the favorite human disease model, the mouse (M. musculus). Traditionally, genetic engineering of mice has primarily been done by reverse genetics. Deletion of a specific gene has been done through generation of knockout mice by homologous recombination, while overexpression of a specific protein has been achieved through the generation of transgenic mice. Many recent technical improvements of these methods, including the usage of genetic recombination systems, such as Cre-loxP, have enabled the generation of tissue-specific and conditional knockout mice, tissue-specific transgenic mice, etc. Different attempts to identify genes involved in biological processes in the mouse in a high throughput manner include random mutagenesis by N-ethyl-N-nitrosourea (ENU), which induces spontaneous mutations in the mouse germline; insertional (gene trap) mutagenesis technology; and, high throughput methods of mutating a large number of genes, as described in Baker et al.,
Amongst these various methods, only the ENU mutagenesis screen is a forward genetics approach, and it is severely limited by the technical challenges in identifying the genetic alterations in mice with a desirable phenotype. However, despite all recent improvements in standard genetic engineering technologies, these methods still remain cumbersome, expensive, laborious, and time-consuming.

[0013] The power of RNAi and the efficiency of retrovirus-mediated gene-delivery methods have recently been combined to generate “knock down” mice. Specific RNAi have been delivered through the mouse germline by transduction of embryonic stem cells with short hairpin (sh)RNA-expressing lentiviruses resulting in stable knock-down of the target proteins in the resulting mouse offspring (Rubinson et al., 2003). The RNAi-based knock down offspring phenocopy the effects of gene knock out by homologous recombination (Kunath et al., 2003). Furthermore, RNAi has also proven to be effective at generating hypomorphic alleles and epiallelic series of gene expression (Hemann et al., 2003), which may more accurately model the changes in gene expression that underlie the development of a number of human diseases, as compared to a complete null allele. Although RNAi is a useful technique for performing high throughput reverse genetics in mammalian tissue culture systems, the technology has not been used to perform in vivo drug target identification, and even more importantly, drug target validation, in mammals in a disease-relevant context.

SUMMARY OF THE INVENTION

[0014] The present invention is related to a method of identifying a gene that encodes a modulator for a biological pathway, comprising introducing to a first mammalian model system of a biological pathway a population of different DNA molecules encoding a population of dsRNA that are substantially identical to at least a region of a population of genes. A DNA is identified from the population of DNA molecules that encodes a dsRNA that modulates the expression of a target gene of the biological pathway, whereby a target DNA is prevalidated. The prevalidated DNA may optionally be introduced to a second mammalian model system, whereby the DNA is further prevalidated by modulating expression of the target gene in the second mammalian model system. The prevalidated DNA is confirmed as a modulator of the
biological pathway by introducing the prevalidated DNA to a third mammalian model system of a biological pathway.

[0015] The methods of the present invention may be used to identify genes that encode modulators of biological pathways involved in diseases including, but not limited to, cancer, metabolic disorders, like diabetes, obesity, osteoporosis, inflammatory disorders, Rheumatoid arthritis, colitis ulcerosa, and neurological disorders. The modulators may be directly or indirectly involved with the biological pathway.

[0016] The methods of the present invention may be performed using mammalian model systems that are cell line or a mammal. The mammalian model systems of the present invention may also be genetically defined. The mammalian model systems of the present invention may also be a transgenic mammal including, but not limited to, a transgenic mouse.

[0017] The dsRNA of the present invention may be a shRNA. A population of retrovirus may comprise the population of dsRNA. The retrovirus may be a lentivirus.

**BRIEF DESCRIPTION OF THE DRAWINGS**

[0018] Figure 1 demonstrates a flow scheme for the generation of an RNAi target library using a bioinformatics approach.

[0019] Figure 2 demonstrates a platform flow scheme for RNAi-based target identification and validation in vivo.

**DETAILED DESCRIPTION OF THE INVENTION**

[0020] The present invention provides improved methods for identifying molecular mechanisms involved in specific diseases, and a clear understanding of their mechanism-of-action in a disease-relevant, physiological context. The present invention is related to the identification and simultaneous physiological validation of disease-associated target genes in mammalian disease models using RNAi.

[0021] The present invention is directed to a method of identifying a gene that encodes a modulator for a biological pathway or phenotype of interest comprising screening a first mammalian model system with an RNAi-based library to identify and functionally validate target genes that modulate the biological pathway or phenotype. The screen may be performed in a first mammalian model system by introducing a population of different DNA molecules that
encode a population of dsRNA that are substantially identical to at least a region of a population of genes and identifying a DNA that encodes a dsRNA that modulates the expression of a target gene of the biological pathway of interest. The identified DNA may optionally be screened in a second mammalian model system to further validate that the identified DNA is a target. A target gene is confirmed as a modulator and validated physiologically by performing an RNAi-based screen in an additional mammalian model systems comprising a pathway- or phenotype-relevant mammal model. Figure 2 provides a flow scheme for a preferred embodiment of the present invention for RNAi-based target identification and validation in vivo.

[0022] The methods of the present invention are useful in the identification and validation of mammalian disease targets. The use of RNAi libraries in in vivo phenotype-driven screens provides a combined forward and reverse genetics approach allowing the analysis of complex genetic interactions and the efficient functional annotation of targeted genes in mammals. By validating the target gene in a physiological, disease-relevant context, the function of the target gene may be identified at an increased rate and with improved quality without any a priori knowledge of either gene or protein function.

[0023] In addition to RNAi screens, the screens of the present invention may be performed using small molecules and recombinant proteins to enable target deconvolution and pathway mapping of phenotype-modifying agents.

1. Library

[0024] The present invention is related to the use of target-specific RNA interference (RNAi) libraries, which may be used to perform genetic interaction screens in vitro and in vivo in mammalian disease models. The libraries of the present invention comprise dsRNAs. The dsRNAs may be siRNA, miRNA, or preferably shRNA, as well as equivalents thereof. The dsRNA comprises a nucleotide sequence which is substantially identical to the nucleotide sequence of the target gene or a variant thereof or a complementary sequence thereto. By "substantially identical" is meant that the sequence of the dsRNA is at least about 80%-90% identical to 10 or more contiguous nucleotides of the target gene, more preferably at least about 90-95% identical to 15 or more contiguous nucleotides of the target gene and even more preferably at least about 95-99% identical or absolutely identical to 19 or more contiguous nucleotides of the target gene.
[0025] The dsRNA library may be a library of any size. The dsRNA library may target each gene of a genome, or a portion thereof. Preferably, the dsRNA library targets from about 25-10,000 genes, more preferably from about 50-5,000 genes, and most preferably from about 100-1,000 genes.

[0026] One consideration in developing the libraries is the optimal design of dsRNA. In contrast to the long dsRNAs used to silence genes in primitive organisms, which can be selected from any part of the mRNA for a gene and are usually very effective at inducing RNAi, si/shRNAs only target a very small region of a mammalian mRNA and exhibit great differences in their ability to prevent target gene transcription. The “rules” for selecting si/shRNA sequences to silence a particular gene are not yet well established and most current algorithms only correctly predict an effective silencing sequence around 35% of the time. Therefore, in order to ensure an efficient knock down of the majority of targets, the preferable size of the dsRNA libraries are four to five times larger than the number of targeted genes. An alternative strategy is to develop libraries in which the dsRNA are validated to ensure that the dsRNA perform post-translational gene silencing, which may be conducted by methods including, but not limited to, measuring RNA levels, real-time PCR, TaqMan, or similar thereto, and at the protein level (e.g. by western blotting, immunohistochemistry, or other methods). It is also important to avoid off-target effects that may influence the results in the design of the library.

a. Focused Library

[0027] The present invention is also related to a focused library and methods of production thereof. In a preferred embodiment, the dsRNA library is focused to target genes encoding molecular targets of one or more particular pathways considered relevant for a particular disease. A molecular target may be a direct member of the particular pathway or directly related to the disease or indirectly associated therewith. Molecular targets that are indirectly associated with the pathway or disease of interest may be identified by criteria including, but not limited to biological, molecular pathological, bioinformatics methods, and proteomics expression profiling methods. Bioinformatics methods include, but are not limited to, ENTREZ Blast and PubMed searching. Figure 1 demonstrates a representative example for the generation of an RNAi target library using such an approach.

[0028] The genes to be targeted by the focused library may be identified by first identifying one or more phenotypes at the cellular and organism level that are associated with the pathway or
disease of interest. For each identified phenotype, there must either be a molecular link to the phenotype in question or a phenotype or function that is modulated by the disease, a phenotypic link, to the disease in question. As a representative example, if the disease of interest is cancer caused by loss-of-function mutations in the tumor suppressor gene PTEN, the associated phenotypes are deregulated growth, survival, angiogenesis, and metastases. In addition, deregulated Akt kinase activity is a key oncogenic driver of the transformed phenotype. As an additional representative example, if the disease of interest is obesity or diabetes, associated phenotypes and functions, so-called phenotypic links, are glucose uptake, fat and protein metabolism, and lipolysis, amongst others. Key molecular links include insulin, IGF-1, and PI3'K signaling.

[0029] The molecular link to the phenotype in question or a phenotype or function that is modulated by the disease are then used as search criteria for a bioinformatics-based search to produce a first gene list which comprise genes known or suspected to functionally and/or physically ‘interact’, in the broadest sense, with the molecular link or phenotype used for the search. In the case of a key molecular link, for example Akt in PTEN-induced cancers, a comprehensive search is done in ENTREZ against gene symbol (Akt) and all aliases (e.g. RAC-A, protein kinase B, etc) for human and mouse. In addition, other standard methods are included, like consensus sequence, domain search, expression, etc. As a parallel approach, a comprehensive search is done in PubMed, Locus link, and other relevant databases against the same gene symbol and aliases. In addition, PubMed abstracts, Blast, and other accessible database may be searched as needed based on biological key question, biological knowledge about the disease, or other specific factors (case-to-case-specific). All genes coming out of such a comprehensive search may be identified and extracted through a simple text-based search against a reference set comprising all known genes and proteins in the genome (available at the NCBI Locus Link web-site). Based on these parallel approaches, a non-redundant gene list is produced, which may then be ranked in descending order based on the number of references for each gene.

[0030] In situations where only limited molecular information is available about the disease of interest, one can identify targets that might be associated with the disease of interest by their published relationship with a key phenotype associated with that disease. As an example, if the disease of interest is diabetes, one can identify targets in pathways associated with insulin and
IGF-1 receptor signaling, including PI3'K and Akt signaling pathways, or one can include targets that have been associated with altered glucose uptake, fat metabolism and distribution, the latter being examples of phenotypic links associated with diabetes.

[0031] As a control of the quality of the search, accessory genes should be identified that are known or expected to interact with the search gene. Representative examples of such control genes include, but are not limited to, genes coding for heat shock proteins. In addition, different fields may optionally be added to the gene list annotation, enabling a manual, biological, qualifying description of the stringency of the association of the individual genes with the query molecule. Such a process ensures that the genes in the resulting gene list meet biological criteria. Representative examples include, but are not limited to, physical association (Y/N), type of physical association (e.g., endogenous proteins, overexpressed proteins, yeast two-hybrid, etc), transcriptional regulation (direct/indirect transcriptional target), post-translational regulation (e.g., stability, degradation, phosphorylation, etc), and so forth.

[0032] In a preferred embodiment, an additional bioinformatics-based search is then performed by searching in a similar manner each of the members of the first gene list to produce a second gene list which comprises genes that are linked with the members of the first gene list. This type of cross-reference provides for additional genes that are associated with the disease or phenotype in question. The extracted genes may then be ranked in descending order based on the number of individual hits from the first gene list that are referenced by each member of the second gene list. In other words, the top of the second gene list are those genes that each refer to the highest number of individual members of the first gene list.

[0033] The present invention specifically contemplates multiple searches being performed using the results of a previous search. The results of each additional search expand the relationship between genes likely to be associated indirectly with the phenotype or disease in question. Moreover, the searches may be performed in an iterative manner.

[0034] In one embodiment, the focused library comprises gene targets from one or more gene lists discussed above. The gene targets are preferably at the top of a particular gene list, which may provide the most relevant targets.

[0035] In another embodiment, the focused library comprises gene targets from one or more gene lists that are reported in high impact journals. This ensures that genes that were reported few times, but in high impact journals, are included in the pool. Representative examples of high
impact journals include, but are not limited to, Nature, Science and Cell. High impact journals may also be based on Science Citation Index classification and individual criteria relevant for the disease in question. For example, if there is molecular, biochemical and cell biological knowledge about the disease in question, (e.g., an oncogene involved in human cancers), the high impact factor journals include molecular oriented journals. If there is primarily clinical knowledge, and only limited molecular knowledge, about the disease in question (e.g., urinary incontinencia), the high impact factor journals include high-ranked clinical journals.

[0036] In another embodiment, the focused library comprises gene targets from one or more gene lists that are cited a minimum number of times and within a recent minimum amount of time. This ensures the presence in the library of gene targets that have been recently discovered or studied, yet are relevant for the disease in question. The number of citations may be from about 1 to about 10, preferably from about 3 to about 5. The period of time may be from about 3 months to about 2 years, preferably about 1 year.

[0037] In another embodiment, the focused library of the present invention comprises a combination of the individual focused libraries discussed above.

[0038] In a preferred embodiment, the individual members of the focused library are filtered based on their drugability, which may be determined according to generally accepted criteria. As a representative example, if the focus is on small molecule inhibitor development, the genes considered most drugable may be enzymes, including kinases, phosphatases, receptors, and ligands. Alternatively, if the focus is on antibody development, the target genes considered most drugable may be based on accessibility to such drugs.

[0039] In another preferred embodiment a relational display is prepared using the key molecular link and/or phenotypic function with the first generation targets surrounding it. The gene list is graphically displayed around the center molecular and/or phenotypic link for quick visual assessment. The relational display may be expanded to repeat the graphical display for one or members of other gene lists, where the resulting list is displayed in a root fashion around the key search molecule. The ultimate result is a relational display of molecular members of an interrelated focused library which is fully integrated.
2. Delivery of dsRNA to Mammalian Model System

[0040] A dsRNA of the present invention is capable of modifying the expression of a target gene in a cell, tissue or organ. Expression of the target gene may be delayed, repressed or otherwise reduced in an animal cell, which expresses the dsRNA.

[0041] dsRNA may be chemically synthesized and transfected into cells. Preferably, dsRNA are generated in vivo from expressed short hairpin RNA (shRNA) precursors, which possess a stem-loop structure, reminiscent of endogenously present micro RNAs (miRNAs). These shRNAs may be coded by 50-70 bp long DNA sequences that are operatively linked to a promoter. The DNA sequence encoding the shRNA may be operatively linked to the promoter in the sense or antisense orientation.

[0042] Reference herein to a "promoter" is to be taken in its broadest context and includes the transcriptional regulatory sequences of a classical genomic gene, including the TATA box which is required for accurate transcription initiation in eukaryotic cells, with or without a CCAAT box sequence and additional regulatory elements (i.e. upstream activating sequences, enhancers and silencers). A promoter is usually, but not necessarily, positioned upstream or 5', of the DNA sequence encoding the dsRNA. Furthermore, the regulatory elements comprising a promoter are usually positioned within 2 kb of the start site of transcription of the the DNA sequence encoding the dsRNA.

[0043] The promoter may regulate the expression of the dsRNA constitutively, or differentially with respect to cell, the tissue or organ in which expression occurs or, with respect to the developmental stage at which expression occurs, or in response to external stimuli including, but not limited to, physiological stresses, pathogens, metal ions, and other inducing agents, like tetracycline, ecdysone, mifepristone, U286, and others. Preferably, the promoter is capable of regulating expression of the dsRNA in a eukaryotic cell, tissue or organ, at least during the period of time over which the target gene is expressed therein and more preferably also immediately preceding the commencement of detectable expression of the target gene in said cell, tissue or organ.

[0044] Strong constitutive promoters are particularly preferred for the purposes of the present invention or promoters which may be induced by virus infection or at the commencement of target gene expression. Examples of preferred promoters include, but are not limited to, the bacteriophage T7 promoter, bacteriophage T3 promoter, SP6 promoter, lac operator-promoter,
tac promoter, SV40 late promoter, SV40 early promoter, RSV-LTR promoter, and CMV IE promoter. Those skilled in the art will readily be aware of additional promoter sequences other than those specifically described.


[0046] Many primary cells are refractory to transfection, but can be readily transduced by viruses. As a consequence, viral vectors are the tools of choice to deliver dsRNAs in order to study gene function in many mammalian cell types. The most popular of these viral systems are based on adenoviruses and retroviruses. Adenoviruses can infect a broad-range of cell-types without the need of active cell division and can be produced at a very high titer. Replication-defective adenoviral vectors have been engineered to efficiently silence gene expression in various cell lines and organs like brain and liver. Most of these vectors have been created by inserting a PolIII promoter and sequences for shRNAs into the genomes of replication-defective adenoviruses. Adenovirus-based RNAi vectors do have some limitations: first, they do not integrate into the genome and therefore provide shRNA expression only in a transient manner in dividing cells and second, they induce a strong immune response that leads to the elimination of the transduced cells.

[0047] Retroviruses are able to integrate into the genome, do not elicit an antiviral response, and are able to transduce a wide range of cell types. Additionally, lentiviruses, a sub-type of retroviruses, can infect non-cycling and post-mitotic cells. Consequently, an important advantage of lentiviral vectors over other viral delivery systems for shRNAs is that they can stably transduce both adult and embryonic stem cells and single cell embryos. This strategy has recently been exploited for the efficient generation of transgenic mice and rats; therefore,
lentivirus-based vectors may be used more generally to generate “knockdown” transgenic
animals that express shRNAs in certain tissues or in all the cells of the animal.

[0048] Replication-incompetent, self-inactivating (SIN) lentiviral vectors may be used in the
practice of the invention. Representative examples of such vectors are discussed in De Palma
and Naldini, 2002; Follenzi and Naldini, 2002. A preferred vector, such as LentiLox 3.7, may
efficiently infect various types of stem cells, immune cells and neurons and can be used for the
generation of such “knockdown” animals. See Rubinson et al., Nat Genet 2003, 33:401-406, the
contents of which are incorporated by reference. The vector may express an shRNA under the
control of the mouse U6 promoter. The vector may also comprise the gene encoding the
Enhanced Green Fluorescent Protein (EGFP) driven by the CMV promoter as a marker to track
infected cells. By using the vectors of the present invention, gene expression may be silenced in
immune cells and hematopoietic stem cells. Moreover, transgenic animals derived from
embryonic stem cells infected with the virus may show markedly reduced expression of the
targeted gene (CD8) in cells present in the central and peripheral immune organs. Similarly,
infecion of zygotes with the vectors may result in efficient silencing of various target genes in
the developing embryo with the effect also being maintained in the adult animal.

[0049] In one embodiment, the encoded short dsRNAs are cloned into viruses as stem loop
dNAs downstream of a Pol II-based U6 or H1 promoter, or a classical Pol III promoter to drive
expression of shRNAs. In a representative example for the viral delivery of RNAi to target cells,
the shRNA itself contains 19 nt-long, complementary sequences, and a short stem loop. These
shRNAs may be enzymatically processed within the target cells, resulting in generation of the
specific RNAi. In addition, the viruses may contain a marker, such as GFP. If needed, the viruses
may be rendered bifunctional, by co-expression or fusion with a sensitizer gene, which is a
disease-relevant gene implicated in the disease-associated signaling pathways including, but not
limited to an oncogene. The viruses can also be rendered inducible, so that they may either have
inducible siRNA expression and/or inducible gene expression (Czauderna et al., 2003; Gupta et
al., 2004). Finally, the viruses may carry recombination markers, such as Cre-LoxP enabling the
excision or inversion of the intervening gene sequence(s) upon successful genomic integration.

The lentiviruses may be expressed and produced in cells, such as HEK-293 cells, according to
standard methods. The resulting lentiviral supernatants may be used individually or as pools
consisting of shRNAs directed against one or more genes.
3. Mammalian Model System

[0050] The mammalian model systems may be tailored to a specific molecular pathway defect or a specific therapeutic area in question. The mammalian model system may be a cell line or intact mammal comprising altered expression of a sensitizer gene. The sensitizer gene is a gene that can be mutated, or has a product with altered expression, function, and/or activity, that results in a phenotype relevant for the disease or phenotype of interest. Examples include, but are not limited to, oncogenes, tumor suppressor genes, differentiation genes, survival genes, genes involved in cell cycle progression, metabolic functions, protein degradation, protein stability, protein turnover, and in control of cell migration, metastasis, etc. Expression of the altered sensitizer gene may result in abnormal proliferation of the target tissue, in other in enhanced cell survival, altered differentiation, cell death, altered cell growth control, adhesion, migration, etc. Through the knock down of the individual target genes for the RNAi libraries, the sensitizer gene function is systematically screened for genes that modulate the sensitizer gene function.

[0051] The mammalian model system may be a cell line. The mammalian model system may also be a multicellular organism including, but not limited to, Drosophila, C. elegans, Xenopus, rat, chicken, zebrafish, and preferably mice. The mammalian model systems useful in the practice of the invention may be based on specific, molecularly defined disease models chosen to represent the corresponding human disease as closely as possible. The use of mammalian model systems according the methods of the present invention will enable overcoming the two major obstacles in drug discovery: lack of efficacy and mechanism-based toxicity in phase II and/or phase III trials.

[0052] Preferably, the mammalian model system is genetically defined. A "genetically defined" mammalian model system refers to any cell line or intact mammal containing one or more cells bearing genetic information altered or received, directly or indirectly, by deliberate and specific genetic manipulation, such as by targeted recombination or microinjection or infection with recombinant virus, and encompasses cells in animals that are altered by, or receive, a recombinant DNA molecule. The recombinant DNA molecule may be specifically targeted to a defined genetic locus, may be randomly integrated within a chromosome, or it may be extrachromosomally replicating DNA.

[0053] The altered expression of the sensitizer gene as well as the knock down of the target genes, may be targeted to the whole animal, or one or more specific and identifiable tissues, or
portions thereof, dispensable for the survival and reproduction of the animal. Such tissue is referred to herein as "target tissue". Examples of preferred target tissues include the ears and certain skin and blood cells, the liver, heart, reproductive organs, brain, and other organs in the mouse. In one preferred embodiment, altered expression of the sensitizer gene and/or the target genes occur in only a portion, patch, sector or domain of the target tissue(s). This can make it easier to identify target genes that modulate the specific function of the sensitizer gene in that portion of the tissue in which the altered expression of the sensitizer gene is manifested. Also, with some tissues, such as the skin, blood, liver, heart, etc., the whole tissue is not dispensable. Thus, altered sensitizer gene or RNAi expression may be targeted to only a portion or sector of the tissue, for example, in melanocytes of skin or T cells of blood, to maintain viability of the animal. The target tissue is one that is identifiable, in that changes in levels of cell proliferation of the tissue can be detected, either visually (e.g. changes in the size of a solid tumor, or the color of skin pigment) or by aided analysis (e.g. analysis of blood cells using a Fluorescent Activated Cell Sorter).


Typically, expression will be under control of tissue- and/or developmental stage-specific regulatory elements (e.g. enhancers and/or promoters). Examples of regulatory elements in the mouse include melanocyte-specific tyrosine-related protein 1 (TRP-1) promoter, which drives expression in the skin (Hart, Semin Oncol (1996) 23(1):154-158; Carreira et al., Mol. Cell Biol (1998) 18(9):5099-5108), the BMP5 promoter (DiLeone et al., Genetics (1998) 148(1):401-408) which drives expression in ear cartilage, and the Lck promoter which drives expression in T-cells (Wildin et al., J Immunol (1995) 155(3): 1286-1295). Other promoters/enhancers are well-known
in the art for limiting expression of a transgene in a cell- or tissue-specific manner in a model organism, and can be used.

[0055] In a preferred embodiment, detection of the expression of the sensitization gene and/or detection of the modulatory effects achieved through the RNAi-directed downregulation of interactor genes is facilitated by the presence or absence of marker gene expression in the target tissue. Any gene can be used as a marker that causes a reliable and easily scored phenotypic change in transgenic animals.

[0056] As a representative example, tumor formation in nude or irradiated syngeneic mice by oncogene-overexpressing, transplanted hematopoietic stem cells can be a prescreen assay for cancers, in which the particular oncogene is involved in the pathogenesis of the human cancer in question. Insulin-controlled glucose uptake in adipose cells or adipocytes in culture could be a prescreen assay for diabetes, neuronal survival or neurite extension could be a prescreen assay for specific neuronal diseases, luciferase reporter assays for PI3K-Akt pathway activity could be a prescreen assay for PTEN-induced cancers. The prescreen assays may also be more tailored by using genetically defined cells or cell lines. The more physiologically relevant the assay in its ability to reflect a key pathogenic element of the human disease in question, the better.

[0057] In one embodiment of the invention, the sensitization gene is a cell-cycle progression gene including, but not limited to, p53, Rb, p27, p15, p16, Bub-1, LATS, CyclinE, E2F, and genes encoding cyclin-dependent kinases. Preferably, the sensitization gene has been identified previously as an oncogene or tumor suppressor gene of the metazoan animal used in the screen, or is a homolog or ortholog of a human oncogene or tumor suppressor gene. Examples of human oncogenes of which homologs have been identified in one or more model organisms include Akt, Abl-1, Bcl2, Ras, CDC25A, several cyclin-dependent kinases, cyclins, serine/threonine kinases, and more than half of the human tyrosine kinase (see (Blume-Jensen and Hunter, 2001; Hanahan and Weinberg, 2000; Hunter, 1997; Rangarajan and Weinberg, 2003; Vivanco and Sawyers, 2002) for recent, comprehensive reviews) Examples of human tumor suppressor genes of which homologs have been identified in one or more model organisms include, and is not limited to, BRCA1, Rb, p16, p53, VHL, and Beta-Catenin. A comprehensive list of oncogenes and tumor suppressor genes may be found at the website for the Cancer Genome Anatomy Project (CGAP) of the National Cancer Institute:
The present invention may be used for the identification and validation of targets for cancers of any type, including solid tumors and leukemias, including, but not limited to all major carcinomas, sarcomas, and hematopoietic malignancies. Examples include: apudoma, choristoma, branchioma, malignant carcinoid syndrome, carcinoid heart disease, carcinoma (e.g., Walker, basal cell, basosquamous, Brown-Pearce, ductal, Ehrlich tumor, in situ, Krebs 2, Merkel cell, mucinous, non-small cell lung, oat cell, papillary, scirrhouss, bronchiolar, bronchogenic, squamous cell, and transitional cell), histiocytic disorders, leukemia (e.g., B cell, mixed cell, null cell, T cell, T-cell chronic, HTLV-II-associated, lymphocytic acute, lymphocytic chronic, mast cell, and myeloid), histiocytosis malignant, Hodgkin disease, immunoproliferative small, non-Hodgkin lymphoma, plasmacytoma, reticuloendotheliosis, melanoma, chondroblastoma, chondroma, chondrosarcoma, fibroma, fibrosarcoma, giant cell tumors, histiocytoma, lipoma, liposarcoma, mesothelioma, myxoma, myxosarcoma, osteoma, osteosarcoma, Ewing sarcoma, synovioma, adenofibroma, adenolymphoma, carcinosarcoma, chordoma, cranio-pharyngioma, dysgerminoma, hamartoma, mesenchymoma, mesonephroma, myosarcoma, ameloblastoma, cementoma, odontoma, teratoma, thymoma, trophoblastic tumor, adenocarcinoma, adenoma, cholangioma, cholesteatoma, cylindroma, cystadenocarcinoma, cystadenoma, granulosa cell tumor, gynandroblastoma, hepatoma, hideradenoma, islet cell tumor, Leydig cell tumor, papilloma, Sertoli cell tumor, theca cell tumor, leiomyoma, leiomyosarcoma, myoblastoma, myoma, myosarcoma, rhabdomyoma, rhabdomyosarcoma, ependymoma, ganglieneuroma, glioma, medulloblastoma, meningioma, neurilemmoma, neuroblastoma, neuroepithelioma, neurofibroma, neuroma, paraganglioma, paraganglioma nonchromaffin, angiookeratoma, angiolymphoid hyperplasia with cosinophilia, angioma sclerosing, angiomatosis, glomangioma, hemangioendothelioma, hemangioma, hemangiopericytoma, hemangiosarcoma, lymphangioma, lymphangiomyoma, lymphangiosarcoma, pinealoma, carcinosarcoma, chondrosarcoma, cystosarcoma phyllodes, fibrosarcoma, hemangiosarcoma, leiomyosarcoma, leukosarcoma, liposarcoma, lymphangiosarcoma, myosarcoma, myxosarcoma, ovarian carcinoma, rhabdomyosarcoma, sarcoma (e.g., Ewing, experimental, Kaposi, and mast cell), neoplasms (e.g., bone, breast, digestive system, colorectal, liver, pancreatic, pituitary, testicular, orbital, head and neck, central nervous system, acoustic, pelvic, respiratory tract, and urogenital), neurofibromatosis, and cervical dysplasia, and for treatment of other conditions in which cells have become immortalized or transformed.
4. Identification and Functional Prevalidation of a Target Gene

[0059] RNAi libraries are initially screened in a mammalian model system that allows identification and prevalidation of target genes in an in vivo context in a, preferably, high throughput manner. The initial screens may be tailored to the specific molecular pathway or specific therapeutic area in question.

[0060] Individual RNAi species, preferably validated, or pools thereof are screened in a first mammalian model system. The screen is preferably performed by initially infecting individual cells, tissue(s), or body part(s) of the model system with a retroviral-based RNAi library, preferably lentiviral-based. By monitoring the phenotype and the kinetics of the phenotypic changes, viruses or virus pools may be identified and functionally prevalidated as positive or negative modulators of the phenotype, or the kinetics of the phenotype, in question. The target genes may be identified from the functionally prevalidated viruses by methods including, but not limited to, PCR amplification of the genomic DNA from the infected target cells using primers against conserved regions in the viruses.

[0061] The identified dsRNAs involved in the phenotype modulation may optionally be verified by infecting a mammalian model system, preferably the same system, with the individual dsRNAs expressed from the corresponding individual viral supernatants. This verification step may be used to identify the role of individual dsRNAs for the phenotype involved. The dsRNAs may cause varying levels of target gene silencing, thereby providing a series of hypomorphic phenotypes that may further reveal the role of the target gene in the specific phenotype and/or disease process.

5. Additional Functional Prevalidation of a Target Gene

[0062] Target genes identified and functionally prevalidated as drug targets through the above methods may optionally be further functionally prevalidated by performing an additional screen of identified dsRNA in an additional mammalian model system. Preferably, identified shRNA-expressing lentiviruses are used to knock down the targets in primary cells, cell lines, and other model systems with different phenotypic read-outs. Overexpression of the targets may also be performed, and testing of their role in specific phenotypic assays, in accordance with their suspected role in pathways, and, finally, testing of their expression in normal and diseased tissues. The further functional prevalidation of the targets identified in the initial screens establishes the targets as important for disease progression and disease progression modulation.
6. Final Confirmation and Physiological Validation of a Target Gene

[0063] Identified and functionally prevalidated targets that modulate the disease phenotype may be confirmed and physiologically validated by screening the relevant dsRNA in a mammalian model system comprising an intact mammal that is model of the phenotype in question. Final confirmation of a target gene demonstrates the ability of the target gene to modulate disease progression in a relevant disease animal model that, preferably, mimics the human disease in question.

[0064] Preferably, shRNA-expressing lentiviral vectors are used to knock down the targets through the germline of mammals, primarily mice. This may be accomplished through direct transduction of either cultured embryonic stem cells, the zygote through injection, the morula, blastula, or even into specific tissues/organs in growing or adult mammals. The use of RNAi to produce "knock-down" mice is described in Carmell et al., "Germline transmission of RNAi in mice," Nat Struct Biol 10:91-92, 2003 and Rubinson et al., "A lentivirus-based system to functionally silence genes in primary mammalian cells, stem cells and transgenic mice by RNA interference," Nat Genet 33:401-406, 2003, the contents of which are incorporated by reference.

[0065] shRNA-expressing lentiviral vectors are preferably made as inducible constructs to circumvent embryonic lethality, by allowing a gene essential for embryonic development to be expressed during embryogenesis, and then knock it down subsequently in the adult animal. The constructs may also be rendered tissue-specific by having the dsRNA expressed under the control of a tissue-specific promoter. The resulting chimeric offspring may then be crossed to yield targeted knock-down offspring. The targeted knock-down offspring may then be crossed with a pathway- or phenotype-relevant mammal. The resulting offspring may be used to finally confirm and physiologically validate the target gene.

[0066] While traditional gene targeting technology requires tedious isogenic genomic DNA cloning, generation of complicated cloning and targeting vector constructs, transfection and homologous recombination of ES cells, ES cell clone picking, retesting for correct insertion of targeting construct by Southern blotting and PCR, the present method requires only infection with the lentiviral vector preparations, and selection of the ES cells by identification of cells where the target gene mRNA and protein expression is knocked down. This can easily be accomplished, since the GFP or other marker in the lentiviral constructs enables easy microscopic identification of the transduced cells, and standard techniques, like RT-PCR,
AmpliTaq-based methods, or western blotting allow for easy identification of the ES cells where the target gene is downregulated. Another advantage with the RNAi-based method, as mentioned above, is that it allows one to generate knock down mice where the target protein expression is only partially reduced, a so-called hypomorphic allele. This has advantages in the validation of the target as a suitable drug target for small molecule or other intervention, where the effects are often only partial.

[0067] In addition, transgenic overexpressors of the same targets may be produced to test their role when overactive. This may be done by cloning the target cDNAs into the lentiviral vector backbone, followed by ES cell infection, and or pronuclear injection of the zygote.

7. Identification of Compounds that Modulate Target Gene Function

[0068] The libraries may also be used to produce genetically defined cell lines, which may be used to perform forward chemical genomics cell-based screens and cell-based screens on libraries of genetically defined cells that are derived from the same naïve cell. This may allow subtraction of compounds or molecules that cause unspecific or general toxicity, or non mechanism-based effects. In addition, this may allow improved target deconvolution, and definition of the relevant disease pathway defects. All results from using the libraries may allow for annotation of overlapping vector sets in totally different areas. By cross-referencing, targets or target ideas in seemingly unrelated disease settings may be identified.

[0069] The present invention contains multiple aspects, illustrated by the following non-limiting examples.

**EXAMPLES**

**Example 1**

**Generation of a Target shRNAi Library**

[0070] A directed shRNA library is created based on a list of growth and apoptosis regulating genes that are known, or are predicted to be, members of cytokine-induced growth and apoptosis signaling pathways in hematopoietic cells. To increase the probability of efficient gene silencing by RNAi, several different shRNAs can be utilized for each gene. Each RNAi targeting sequence is chosen based on established rules for identifying efficient silencing sequences.
Example 2
Validation of the shRNAi Library

[0071] Different methods to validate the individual shRNAs are utilized based on either validating the expression level of target mRNA transcripts or target protein after transduction of relevant cells with the specific lentiviral preparations. Validation is performed using real-time PCR of endogenous mRNA levels from a relevant cell type, real-time PCR of transgene-transcribed mRNA encoding the target protein in cells, western blotting detection of either endogenous or overexpressed target protein in cells.

Example 3
Creation Of A Library Of Bi-Functional Lentiviral Vectors

[0072] A bi-functional lentiviral vector is constructed to express the oncogene sensitizer gene Akt and a library of shRNAs. Briefly, a lentivirus RNAi vector is modified to drive expression of the sensitizer gene under the control of the constitutive CMV enhancer-β-actin promoter (CAG).

[0073] Oligonucleotides encoding each of the validated shRNAs of Example 1 are synthesized and cloned into the vector under the control of the U6 promoter. Each construct is validated by restriction mapping and sequencing.

[0074] This library of vectors is distributed in pools, each consisting of about 20 vectors. Each pool contains all four targeting constructs against either four or five distinct genes. An additional pool contains vectors that expressed no shRNAs or shRNAs targeting species-unrelated control genes, including Green Fluorescent Protein (GFP). In an attempt to avoid redundant or compensatory effects by related genes, only one member of a specific gene family is targeted in each pool. High titer lentiviruses are generated by transient transfection of 293T packaging cells and subsequent ultracentrifugation of the supernatant.

Example 4
Effect of RNAi on Tumorigenesis

[0075] High titer lentiviruses produced from the vectors present in the pools in Example 3 are used to infect hematopoietic or other stem cells, which are then administered to lethally
irradiated mice. Hematopoietic or other tumor development is monitored on a daily-basis. Upon
detection of tumors, or degradation of health status, mice are sacrificed and the thymus, spleen
and bone marrow, as well as other organs showing evidence of tumor growth, are harvested.
Tissue samples are analyzed by histology and, in the case of hematopoietic stem cells,
phenotyped for markers of differentiated hematopoietic lineages

[0076] Mice that show accelerated tumorigenesis indicate that the genes targeted by the
Corresponding shRNAs included negative regulators of the sensitizer gene. By contrast, mice
That show delayed tumorigenesis indicate that the genes are targeted by shRNAs that are positive
Regulators of sensitizer-induced oncogenesis. The results from this initial screen functionally
Prevalidate the target genes in an *in vivo* context.

**Example 5**

**Effect of RNAi on Cell Lines**

[0077] To further functionally validate the genes coming out of screen in Example 4, the
corresponding shRNA-containing lentiviruses are used to knock down the cognate target genes
in selected cancer cell lines, as well as primary human cancer cells. This is done for both the
shRNA-containing viruses alone, as well as for the vectors expressing both the shRNA and the
oncogene in question, for instance Akt. A control for the latter is expression of Akt alone with
mock shRNA. The induced cells and cell lines are then examined for their resistance to apoptotic
stimuli, serum starvation, chemotherapy, UV light, and other insults in proliferation, survival,
cell mass, cell cycle progression, and other assays. Based on these results, together with the
results from Example 4, a number of targets are identified and functionally validated for final
confirmation in a mammalian disease context.

**Example 6**

**Effect of RNAi in a Mammalian Disease Model**

[0078] The target genes that are functionally validated as cancer targets in Example 4 and
Example 5 are now ultimately confirmed in a disease mouse model of the original human cancer
in question. In the case of Akt, PTEN loss-of-function (LOF) mutant mice are the disease mouse
model of interest. Using the shRNA-containing lentiviral vectors with an inducible promoter, and
without co-expression of any oncogene, but with co-expression of a marker gene, for instance GFP, embryonic stem cells are transduced with the specific lentiviral supernatants that cause knock down of the target gene in question, or zygotes are injected with these lentiviral supernatants. The inducibility of the constructs ensures that one can circumvent embryonic lethality and generate mice through the germline, where the target protein encoded by the target gene can be downregulated, or even completely knocked down, in the adult animal.

[0079] Once the mice have been generated, they are crossed with the disease mouse model in question, for instance the PTEN LOF mutant mouse. The tumor progression rate for different tumors in the crossed mouse offspring is compared to the tumor progression rate of PTEN mutant mice alone. This can be done invasively, or, preferably, non-invasively, using e.g. MRI, PET, CT scanning etc. If the PTEN LOF mutant mouse crossed with a specific target knock down has significantly delayed tumor progression in one or more tissues, and/or altered tumor spectrum and penetrance, this shows that modulation of the specific target causes an anti-cancer effect \textit{in vivo} in a real disease setting.
CLAIMS

1. A method of identifying a gene that encodes a modulator for a biological pathway, comprising:

(a) providing a population of different DNA molecules that encode a population of dsRNA that are substantially identical to at least a region of a population of genes;

(b) introducing said population of DNA to a first mammalian model system of a biological pathway;

(c) identifying a DNA from the population of DNA molecules that encodes a dsRNA that modulates the expression of a target gene of said biological pathway of said first mammalian model system, thereby prevalidating a target DNA;

(d) optionally performing (a)-(c) in a second mammalian model system with one or more prevalidated DNA identified in step (c), whereby a DNA is further prevalidated by modulating expression of said target gene in said second mammalian model system;

(e) confirming that the prevalidated DNA modulates said biological pathway by:

(i) introducing said prevalidated DNA to a third mammalian model system of a biological pathway; and

(ii) identifying a DNA that encodes a dsRNA that modulates the expression of a target gene of said biological pathway of said third mammalian model system, thereby confirming a DNA that modulates said biological pathway.
1. **ENTREZ search and PubMed search for relationship with key target or key phenotypic link**

2. **GENELIST I: Non-redundant gene list:**
   ranking in descending order based on # references associated with each gene

3. **GENELIST II: Non-redundant gene list:**
   Which genes are cross-referenced with genes from list I?
   - Extraction of all individual genes cross-referenced in abstracts for each gene from GENELIST I
   - Ranking in descending order based on # of individual genes from GENELIST I referenced for each gene

4. **SUBLIBRARY A:** the top of ranked GENELIST II. Manual intervention/control of a fraction of genes

5. **SUBLIBRARY B:** All genes from GENELIST I (and II) with <3 references AND in top impact factor journals

6. **SUBLIBRARY C:** All genes from GENELIST I with <3 references and publ. date < 1 year => manual curation.

7. **FINAL TARGET LIBRARY:** A+B+C.
FIGURE 2

Step I, HT screen: hematopoietic precursor screen in transplanted mice
- 2-300 drugable genes screened for modulation of tumor progression rate

Functionally pre-validated cancer target genes identified

Step II: RNAi knock down in mice of identified cancer genes
F1 offspring crossed into relevant cancer mouse model

Confirmed and physiologically validated cancer target genes
(in vivo disease context)

Optional additional functional validation in human cancer cells and tissues