INNOVATION

New tools for functional mammalian cancer genetics

Thijn R. Brummelkamp and René Bernards

Knowledge of the function of individual genes that encode components of cell-signalling pathways is crucial to our understanding of normal growth control and its deregulation in cancer, but we have functional information for only ~15% of human genes at present. Several new technologies have recently become available to identify gene function in mammalian cells using high-throughput genetic screens. These new tools will make it possible to identify new and innovative classes of anticancer drugs, including those that show synthetic lethal interactions with cancer-specific mutations.

Cells derived from cancerous lesions often resemble their normal counterparts, but differ through the presence of a series of genetic alterations. These alterations are combinations of activating mutations in oncogenes and loss-of-function mutations in tumour-suppressor genes (for a review, see REF. 1). Other than these well-studied genetic alterations, cancer cells possess a number of phenotypic alterations that do not have a known genetic basis. Tumour-cell immortality, for instance, is often induced by activation of the telomerase catalytic subunit, which disables a counting mechanism that limits the number of replications that a normal cell can undergo[2]. However, the genetic alterations that lead to the activation of the telomerase gene are largely unknown and it is a complex task to identify the genes that are responsible for this activation. Insight into the genetic alterations that cause phenotypes such as telomerase activation will allow for the development of new therapeutic intervention strategies. One of the ways in which these insights can be obtained is through functional genetic approaches.

Functional genetic screens in model organisms, such as *Drosophila* and *Caenorhabditis elegans*, have provided valuable information on components of mammalian signal-transduction pathways; however, such model organisms also have limitations for cancer research. First, not all human tumour-suppressor genes are conserved in lower eukaryotes. This is perhaps not surprising, as the increased complexity of mammals — both in number of cells and in number of cell types — might require a far more sophisticated machinery to control faithful DNA replication and cell differentiation than low-complexity and short-lived model organisms. Because sequence analysis of the *C. elegans* genome has not revealed the *MYC* oncogene, and not all of the components of the ARF–MDM2–p53 pathway are present in worms or flies, processes that are regulated by such genes cannot be studied in these organisms. Furthermore, control over the number of divisions a cell can undergo by regulated expression of the telomerase catalytic subunit is not seen in mice, indicating that even mouse cells are an inappropriate model to study cellular ageing by telomere shortening. Indeed, first insights into the regulation of telomerase expression were recently obtained in a genetic screen in human cells[3]. An even more complicated task would be the development of genetic screens to study processes that are important for tumour angiogenesis or metastasis (such as cell adhesion, cell migration or oxygen sensing) in the genetic model systems that are available at present.

The use of mammalian cells to study the genetics of cancer development has been hampered by the lack of effective genetic tools that can be applied in these cells. In the past few years, our arsenal of genetic tools for functional mammalian cancer genetics has expanded markedly. So, what are these genetic tools and what successes have been achieved with their use in cancer research?

Gain-of-function genetic screens

An alteration in cellular phenotype can be accomplished through the introduction of exogenous genetic material into cells. Such an altered phenotype can often be readily selected for in a population of cells. It is most straightforward to select for a phenotype that confers a new trait to cells, such as the ability to form tumours in experimental animals, growth in semi-solid media or acquisition of a transformed morphology. In each case, it is important that the frequency with which the recipient cells spontaneously acquire the selected phenotype is very low. In fact, engineering the proper cell system to carry out a genetic screen is often far more time-consuming and demanding than the performance of the genetic screen itself. As the novel phenotype is introduced in a dominant fashion by the foreign DNA, such genetic screens are referred to as 'gain-of-function' screens. There are various technologies available to introduce foreign genetic material into cells.

DNA transfection. The discovery of methods to introduce foreign DNA into mammalian cells, now some 30 years ago, provided the first powerful tools to study gene function on a large scale[4]. By introducing exogenous DNA into mammalian cells — resulting in integration of the exogenous DNA into the

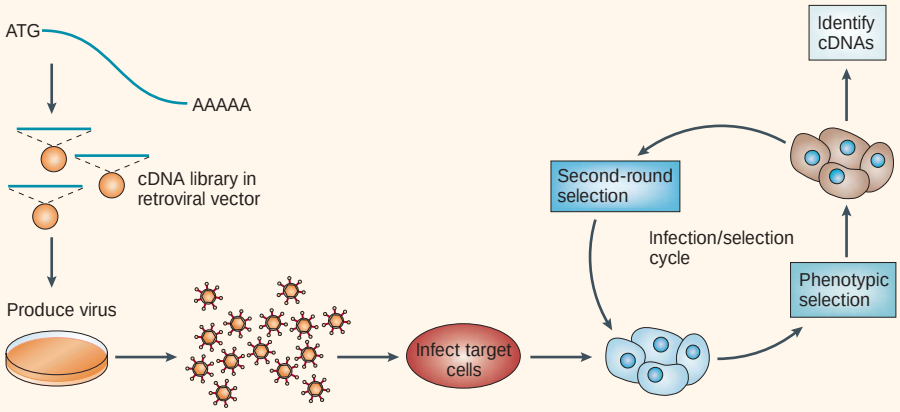


Figure 1 | **Retroviral cDNA library screen.** High-complexity retroviral complementary DNA (cDNA) libraries can be obtained through a single cloning step of cDNAs into a retroviral plasmid vector. High-titre viral supernatants are produced through transfection of the cDNA library vectors into retroviral packaging cells[13]. These viral supernatants are subsequently used to infect the target cells so that all the cDNAs that are present in the library are expressed in the target-cell population. The infected cells can be used for one round of phenotypic selection. After one or more cycles of infection/selection, the cells can be used for virus recovery of the cDNAs that are responsible for the phenotype.

genome of the recipient cell — it became possible to select for cells that had a stable alteration in phenotype. Probably the first gain-of-function genetic screen in mammalian cells that assigned a function to a specific gene was the identification of the fragments of adenovirus type 5 DNA that harbour transforming activity[5]. About 10 years later, the introduction of sheared human tumour DNA into non-transformed NIH-3T3 cells led to the identification of the *RAS* oncogene[6,7].

Complex complementary DNA (cDNA) expression libraries (that is, libraries in which both abundant and rare messenger RNAs (mRNAs) are represented) can be efficiently introduced into mammalian cells using plasmids that replicate episomally (that is, without integrating into the host genome, as they have their own origin of replication). These plasmids have two marked advantages: there is no integration into the host genome — often a rather inefficient process — so there is a higher efficiency of cells that stably express a cDNA; and the vectors can easily be shuttled back into *Escherichia coli* to identify the cDNA that encodes the selected biological activity. For cDNA-library expression in human cells, Epstein–Barr virus (EBV)-based cDNA expression shuttle vectors are used most often[8]. These cDNA libraries have proved to be very useful for cloning genes, especially through complementation of genetic defects in cell function. For example, this cloning strategy has allowed the identification of several Fanconi anaemia genes through functional complementation of the intrinsic sensitivity of Fanconi anaemia cells to mitomycin-C-mediated cell killing[9–12].

Viral vectors. DNA transfection into cells also has its limitations. First, uptake of foreign DNA varies between cell types and is often particularly inefficient in primary (untransformed) cells. The use of such primary cells for genetic screens can be useful — for instance, to identify genes with immortalizing or transforming activity. Second, recovery of the DNA fragment that has the selected biological activity from the transfected cell is both time-consuming and labour-intensive. Both types of problems are bypassed when retroviral vectors are used as a vehicle to introduce foreign DNA into cells.

The first retroviral vectors consisted of simple replication-defective Moloney murine leukaemia virus derivatives, which can be stably maintained after infection, as they integrate into the host genome. High-complexity cDNA expression libraries can be cloned into such retroviral vectors and transfected into 'packaging cell lines'[13] — which produce the viral *gag*, *pol* and *env* genes that are lacking from the replication-defective retroviral vectors — to obtain high-titre retroviral supernatants[14–16] (FIG. 1). These are then used to infect target cells, which leads to stable integration of the provirus into the host genome and expression of the cDNA that is contained by the provirus. These libraries have been used successfully for the identification of cDNAs that induce a selectable phenotypic alteration in the target cells[14–19]. An outline of such a genetic screen is presented in FIG. 1.

When a complex mixture of retrovirus-encoded cDNAs is introduced into a population of cells and a rare variant with an altered phenotype is identified, it is important to verify that the phenotype is caused by the cDNA carried by the retroviral vector in a second round of selection. Because the background of false-positive cells in a genetic screen is often caused by genetic or epigenetic alterations of the target cells, it is mandatory to recover the integrated proviruses from the selected target cells and to perform a second round of selection in a fresh population of target cells (the same applies to genetic screens that use other methods to introduce exogenous DNA into cell populations). When the frequency of cells with the selected phenotype is significantly higher in second-round selection in the infected cell population compared with a control population, this indicates that the selected phenotype is virus-transduced and the selected cells can be used to recover and identify the cDNAs that are responsible for the acquired phenotype. To perform such second-round screens, efficient tools are needed to recover the integrated provirus from the selected cells. Several methods are now routinely used to accomplish this, which greatly facilitate functional cloning strategies that involve retroviral vectors (BOX 1).

High-complexity retroviral cDNA expression libraries have been used to clone transforming oncogenes in NIH-3T3 cells; in one of the first of these screens, a large number of genes with transforming activity was cloned, for example, *RAF1* and *β-catenin*[16] (for additional examples, see TABLE 1).

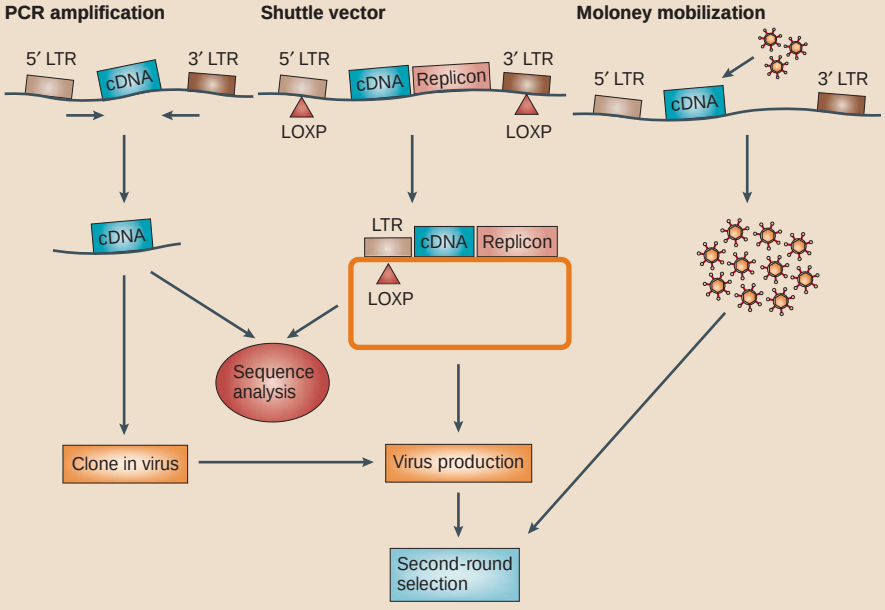
Recently, other viruses have been adapted to harbour exogenous DNA for gene-delivery strategies. For instance, lentiviral vectors have an advantage over Moloney-virus-based vectors as they also infect resting cells, whereas Moloney-virus-based vectors only stably integrate into replicating cells[20].

Replication-defective adenoviral vectors have also been used to generate cDNA expression libraries[21]. Advantages of this vector system are the very broad range of cell types that can be infected by adenoviral vectors (both dividing and non-dividing cells) and the high stability in solution (resistant to several rounds of freeze-thawing) of the adenoviral vectors compared with retroviral vectors.

Arrayed-format libraries. Arrayed-format libraries are collections of individually-cloned cDNAs deposited in a grid-like fashion in multi-well plates. A key advantage of the use of arrayed libraries is that when a desired phenotype is detected in a screen it is directly known which gene caused the phenotype. Whereas genetic screens performed with large polyclonal pools of cDNAs often require repeated rounds of selection to identify the cDNA that confers the selected phenotype, the use of arrayed-format libraries allows the direct

Box 1 | Recovery of integrated retroviral vectors

Several techniques exist for the efficient recovery of the complementary DNAs (cDNAs) that are harboured by integrated proviruses from the genome of a selected cell population (see figure). First, cDNAs can be recovered from genomic DNA isolated from selected cells by polymerase chain reaction (PCR), using primers located in the retroviral-expression cassette that flanks the cDNA insert. These PCR-amplified cDNAs can then be sequence analysed and re-cloned into a retroviral vector[14,16]. When the isolated cDNAs are cloned into viral vectors, new viral supernatants can be obtained by transfection of packaging cells followed by infection of target cells for second-round selection. PCR isolation, however, is more suitable for the isolation of one or a few genes from target cells, rather than for the isolation of more complex mixtures of cDNAs from the target-cell population (because of selective amplification of certain cDNAs). To overcome this limitation, shuttle vectors have been developed, which allow simultaneous recovery of a high number of integrated proviruses from a polyclonal population of target cells. These vectors contain sequences — which are located between the viral long terminal repeats (LTRs) — for maintenance as plasmids in *Escherichia coli* (replicon), so that they are present in the integrated provirus in the infected cell. Sequences that are recognized by site-specific recombinases (such as CRE, which recognizes specific sites (LOXP) in DNA) or restriction enzymes are located in the viral LTRs, so that the proviruses can be excised from the cellular genome by the recombinase or restriction enzyme, circularized and transformed into bacteria[68,69]. Plasmid DNA of a mixture of these proviruses can then be isolated and transfected into packaging cells. Finally, integrated proviruses can be mobilized in a natural way by superinfection of the target cells with wild-type Moloney virus, which complements the provirus by expressing *gag*, *pol* and *env* in the cells[70]. Replication-defective retroviruses that harbour cDNAs will then be packaged into retroviral particles, and viral supernatants that contain both replication-deficient cDNA-encoding viruses and wild-type Moloney viruses can be harvested. This mixture of viruses can than be used to infect a fresh batch of target cells for a second round of selection.



linkage of genotype to phenotype in a single round of screening. This can save an enormous amount of time and effort, and, importantly, also allows selection of phenotypes — such as the induction of apoptosis — that are counter-selected when genetic screens are carried out in a polyclonal format. The stability in solution of the adenoviral vectors greatly facilitates the generation and handling of such arrayed-format cDNA libraries[21] (see BOX 2a). An inherent problem that is associated with the generation and use of cDNA libraries is that some mRNAs are far more abundant than others, leading to unequal representation of cDNAs in a library. Unless measures are taken to prevent this (for example, through 'normalization'; a tedious process that is often carried out at the expense of the average length of cDNAs in the library), large numbers of cDNAs need to be screened to cover the entire complexity of a cDNA library. This is particularly problematic in arrayed-format libraries, as the cDNAs are screened one-by-one for biological activity, which is a very expensive process.

Another ingenious way to perform genetic screens in an arrayed format is to use 'reverse transfection'. In this approach, the transfection is 'reversed'; that is, DNA is not put on top of cells — as is done, for instance, in calcium phosphate transfection — but, rather, the cells are placed on top of spotted-DNA expression vectors. For example, the expression-vector DNA can be spotted onto glass slides, akin to the spotting of DNA microarrays[22]. Cells that are grown on these DNA arrays take up the spotted DNA and can subsequently be used for phenotypic selection (see BOX 2b).

Loss-of-function genetic screens

Genetic screens that aim to identify gene function through inactivation of a gene (or its corresponding mRNA) are referred to as loss-of-function screens. Various technologies have been developed to study the effects of gene suppression in mammalian cells, including genetic suppressor elements (GSEs), antisense vectors, ribozymes, aptamer libraries and, more recently, RNA interference (RNAi). Below, we discuss these methodologies with emphasis on the recent developments in the field of RNAi.

Genetic suppressor elements. GSEs are short and biologically active cDNA fragments that interfere with the function of their cognate gene. GSEs can either encode peptide fragments of proteins that act as dominant-negative inhibitors of the full-length protein or antisense RNA molecules that interfere with the function of the complementary mRNA. When expressed in retroviral vectors, high-complexity GSE libraries can be powerful tools to identify cancer-relevant genes, including determinants of cellular sensitivity to anticancer drugs[23,24] and regulators of transcription[25]. A limitation of this approach is that only a few fragments of any given cDNA act as GSEs, which limits the use of this technology in genome-wide screens. A selection for GSEs with transforming activities resulted in the isolation of ING1 as a regulator of cell proliferation[26].

Antisense vectors. Expression of antisense cDNA molecules can be an effective way to inhibit expression of a certain gene. Libraries that express antisense RNA molecules can be especially useful for studying tumour-suppressor genes (for a review, see

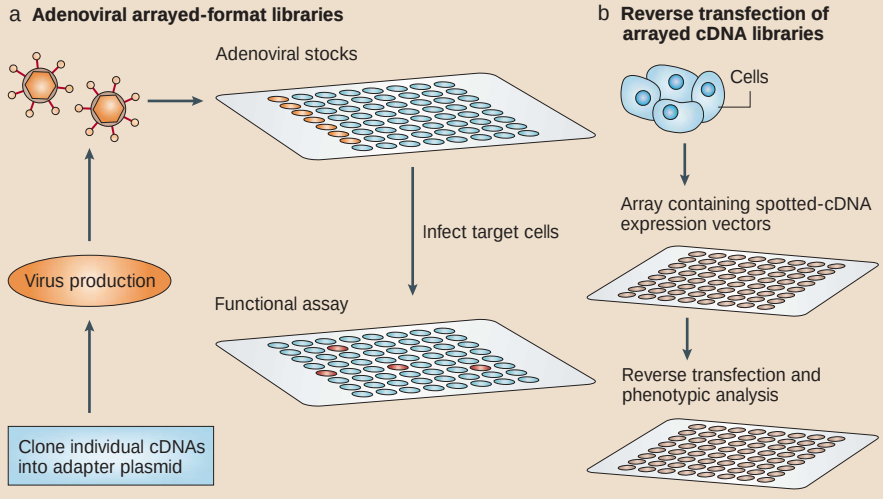
Table 1 | Genes relevant to cancer identified in functional genetic screens

Gene	Type of screen	Phenotypic selection	Reference
HRAS	DNA transfection	Transformation	6
β-catenin	Retroviral cDNA expression library	Transformation	16
MIF	Retroviral cDNA expression library	Inhibits p53-induced cell-cycle arrest	71
Twist/Dermo	Retroviral cDNA expression library	Inhibits c-MYC-induced apoptosis	72
BCL6	Retroviral cDNA expression library	Inhibits p53-dependent senescence	17
FACC	EBNA cDNA expression library	Complementation of genetic defect	9
NEMO	Retroviral cDNA expression library	Complementation of genetic defect	73
DAPK	EBNA antisense library	Resistance against γ-interferon-induced apoptosis	30
CYLD	High-throughput RNAi	NF-κB-regulated reporter	64
TOSO	Retroviral cDNA expression library	Inhibits FAS-induced apoptosis in T cells	19
ING1	GSE screen	Selection for GSEs that promote neoplastic transformation	26
TFE3	Retroviral cDNA expression library	Co-operation with SMAD proteins to stimulate TGF-β-induced transcription	70

BCL6; B-cell non-Hodgkin's lymphoma 6; *CYLD*, cylindromatosis gene; *DAPK*, death-associated-protein kinase; EBNA, Epstein–Barr-virus nuclear antigen; *FACC*, Fanconi's anaemia group C; GSE, genetic suppressor element; *ING1*, inhibitor of growth family, member 1; *MIF*, migration inhibitory factor; *NEMO*, NF-κB essential modifier; NF-κB, nuclear factor κB; RNAi, RNA interference; *TFE3*, transcription factor μE3; TGF-β, transforming growth factor-β.

Box 2 | Arrayed-format libraries

Arrayed-format libraries contain expression vectors that are deposited in an arrayed manner in defined locations. Using this approach, an adenoviral expression library has been generated that contains separated viral stocks that are able to induce expression of individual complementary DNAs (cDNAs; see **a**)[21]. These viral stocks can be used to infect target cells and these cells can be used for phenotypic analysis in a high-throughput manner. Another way to screen arrayed-format libraries is the use of reverse transfection combined with microarrays of spotted-cDNA expression vectors[22] (see **b**). Cells can be transfected (by reverse transfection) on microarrays with individually spotted cDNAs, resulting in spots of locally transfected cells with a certain cDNA. These cells can then be used for phenotypic screens.



REF. 27), and EBNA (Epstein–Barr-virus nuclear antigen) vectors that express an inducible antisense cDNA library have been used successfully to clone several antisense cDNA fragments that protect HeLa cells against γ-interferon-induced apoptosis. The genes that were affected by these antisense cDNAs were named *DAP* (for death-associated proteins) genes[28–30]. As expected, overexpression of *DAP* genes is often incompatible with cell survival[29].

Ribozymes. Another way to inhibit gene expression is by using catalytic RNA molecules — known as ribozymes — that bind to and cleave certain RNA targets. RNA target molecules are recognized through sequence complementarity between the ribozyme and the target RNA molecule. Ribozymes can be effectively expressed in cells using viral vectors, and libraries can be created through the insertion of randomized sequences in the ribozyme RNA sequence. These libraries can be used in genetic screens to identify ribozymes that inhibit certain mRNA molecules. In this way, ID4 has been identified as a regulator of *BRCA1* expression[31] and *PPAN* as a gene that inhibits anchorage-independent proliferation[32]. Also, functional differences between the closely related histone acetyl transferases p300 and CBP were identified[33].

Aptamer libraries. One method that does not depend on cDNA libraries or information on genome sequence is the use of random short-peptide (aptamer) expression libraries. Aptamer expression libraries can be used to find peptides with a desired biological activity[34]. This can be caused by the expressed peptide activating or inhibiting a cellular factor. For stable expression of peptides in mammalian cells, libraries are constructed that express aptamers in the context of protease-resistant scaffold structures. These libraries allow the isolation of peptides that confer cellular resistance to the anticancer drug taxol, presumably through upregulation of multidrug transporters[35]. A limitation of this approach is that highly complex mixtures of aptamers are required to identify active aptamers in any pathway (far more complex than, for example, cDNA libraries). In addition, it can be cumbersome to identify the cellular factor that is affected by the biologically active peptide. In most cases, quite a bit of additional work (for example, a yeast two-hybrid screen) is required to identify the cellular protein that is affected by the interaction with the selected short peptide.

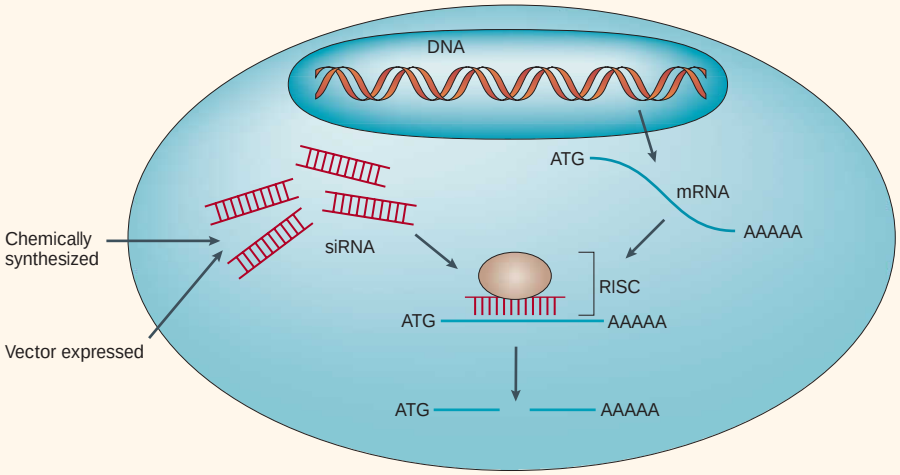


Figure 2 | **RNA interference.** Short double-stranded RNA molecules — short interfering RNAs (siRNAs) — that are introduced into a cell will be incorporated into the RNA-induced silencing complex (RISC). The antisense strand of the siRNA molecule is then used for the identification of cellular messenger RNA (mRNA) molecules that contain a perfect complementary sequence. These mRNA molecules are subsequently cleaved by RISC and become unstable. As a result of this, the protein encoded by the mRNA will not be produced. The siRNAs can be synthetically synthesized and transfected into the cells (this will only allow transient inhibition of gene expression) or they can be produced in the cells by DNA-based vector systems that direct expression of siRNA-like molecules (this can also be used for long-term silencing).

RNA interference. The ability to manipulate gene expression in the multicellular worm *C. elegans* — by techniques such as mutagenesis by transposon tagging, introduction of DNA deletions by chemical mutagenesis and by antisense RNA — made this organism popular among geneticists[36]. The availability of these methods, combined with the short reproductive cycle of *C. elegans*, has made it possible to dissect complex genetic pathways that regulate diverse biological processes. The discovery of a cellular response against double-stranded RNA[37], however, has provided one of the most powerful tools to turn off gene expression, and this has revolutionized loss-of-function genetics, initially in *C. elegans* and *Drosophila* and also more recently in mammalian cells.

This cellular response, known as RNAi, is an ancient defence mechanism to protect cells from foreign invasion by agents such as transposons and viruses (for reviews, see REFS 38,39). The double-stranded RNAs that are produced by integrated transposons or by replicating viruses are processed into short double-stranded RNAs, named short interfering RNAs (siRNAs), that serve as a signal to activate the RNA-induced silencing complex (RISC), which degrades homologous mRNA molecules. The introduction into a worm of double-stranded RNAs that are identical to cellular transcripts will cause degradation of the corresponding mRNA molecules, leading to post-transcriptional gene silencing (FIG. 2).

Genome-wide high-throughput RNAi screens have been performed with success by feeding worms collections of bacteria-containing plasmids that express long double-stranded RNAs that are homologous to a large number of worm genes. Such large-scale genetic loss-of-function screens have allowed the identification of genes that affect worm development[40,41], regulate fat storage[42] or serve to protect the genome against mutations[43]. A similar approach was used in a *Drosophila* cell line to isolate *Hedgehog* pathway components[44]. An advantage of these screens is that it is immediately known which gene is inhibited. By contrast, mutagenesis screens require laborious mapping of the mutation before the loss-of-function phenotype can be linked to a specific gene.

The completion of the sequence of the human genome and the identification of the encoded genes opens the possibility to use RNAi for the systematic inactivation of large numbers of human genes. However, the use of RNAi in mammalian cells is hampered by the fact that the introduction of long double-stranded RNA molecules (which are effective in worms and flies to silence gene expression) leads to nonspecific toxicity in most somatic cells because the γ-interferon pathway is activated. This problem can be solved by the use of chemically synthesized 21–23 base-pair double-stranded siRNAs that are introduced by transfection. These short RNAs can recruit mammalian RISC to target mRNAs and induce their specific degradation, but do not activate the γ-interferon pathway[45]. Sequence-assisted annotation of all human open reading frames (ORFs), also known as the ORFeome, can function as a guide to construct a collection of siRNAs that are designed to inhibit the expression of individual genes.

Although transfection of human cells with siRNAs that are produced *in vitro* is a quick way to silence a gene of interest, there are several problems associated with the use of this method. The first obstacle is the transient nature of the inhibition of gene expression, as the siRNA molecules are unstable inside the cell. Second, several cell types (for example, primary cells) are difficult to transfect with high efficiency. Third, chemically synthesized siRNAs are expensive, which becomes an obstacle when genome-wide RNAi screens are contemplated in mammalian cells.

Several groups have effectively resolved these issues through the development of vector systems that mediate stable production of siRNA-like molecules in mammalian cells[46–50] (FIG. 3). These vectors use RNA-polymerase III promoters to direct the synthesis of short hairpin RNA (shRNA) molecules, which are processed intracellularly into siRNA-like molecules. The shRNAs contain a perfectly double-stranded stem of 19–29 base pairs, which is identical in sequence to the mRNA that is targeted for suppression, connected by a loop of 6–9 bases, which is efficiently removed *in vivo*. Vector-produced shRNA molecules are as effective as siRNAs that are generated *in vitro* in inhibiting gene expression, and, in addition, can be used to study loss-of-function phenotypes that develop over longer periods of time. Furthermore, stable integration of these expression cassettes can be efficiently achieved through retroviral or lentiviral delivery[51–57].

Several observations indicate that vector-mediated RNAi can also be used to study loss-of-function phenotypes *in vivo*. Human cancer cells that stably express siRNAs directed against the *KRAS* oncogene lose their transformed phenotype not only *in vitro*, but also *in vivo* after injection into nude mice[52]. This indicates that gene inhibition by RNAi is maintained in a living animal. Consistent with this, *in vivo* transfection of siRNAs into the livers of mice can inhibit expression of co-transfected reporter plasmids[58]. Transgenic embryos have also been generated that express an shRNA directed against a specific mouse gene, and these mice show the same phenotype as mice carrying a null mutation in the same gene[59]. Together, these studies

indicate that it might be possible to perform large-scale RNAi screens in mice to identify loss-of-function phenotypes that cannot be readily selected *in vitro*.

One potential complication in using siRNAs in genetic screens is that siRNAs are not always completely target-specific. Some 'off-target' side effects have been reported[60–62], which in most cases can be attributed to partial homology of the un-intended target to the siRNA. Even though these off-target effects of RNAi are modest, compared with antisense RNA approaches, for example, it highlights the need to validate an identified phenotype caused by an siRNA with a second, independent siRNA directed against the same transcript. In rare cases, shRNA vectors have been reported to induce an interferon response, the molecular mechanism of which remains unclear[63].

The reduced cost of vector-based RNAi also makes it possible to carry out genome-wide genetic screens for loss-of-function phenotypes in mammalian cells. High-throughput cloning of short-hairpin-encoding DNA oligonucleotides into retroviral vectors containing RNA-polymerase III promoters is now in progress in several places, including our own laboratory. As a proof of concept for the genome-wide RNAi vector library approach, we have recently constructed a 'gene family' knockdown library, which targets most of the members of the family of de-ubiquitylating enzymes. Using this library, we identified the cylindromatosis tumour-suppressor gene (*CYLD*) as a key regulator of NF-κB activity in a high-throughput reporter-based genetic screen[64].

Novel applications of RNAi vectors

One of the more attractive applications of RNAi vector libraries is the identification of 'synthetic lethal' interactions in mammalian cells — a combination of two non-lethal mutations that, together, result in cell death. Such a genetic association indicates that a functional interaction between the corresponding gene products exists. In simple organisms, such as yeast and *C. elegans*, many synthetic lethal interactions have been identified; by contrast, there are few examples of synthetic lethality in mammalian systems. This is probably because the tools to efficiently identify such interactions in mammalian cells have, so far, been lacking. Identification of synthetic lethal interactions might be of key importance for the identification of novel and more powerful classes of cancer drug targets. For instance, inhibition of a gene that shows a synthetic lethal interaction with loss of *TP53* will only be toxic in cells that contain a mutation in *TP53*, making it a tumour-cell-specific drug target. Inhibition of such genes in normal cells (that contain wild-type *TP53*) will be much less toxic, providing a therapeutic window for cancer treatment. The problem in identifying siRNA vectors that show a synthetic lethal interaction with cancer-specific mutations is that these vectors are specifically lost from cells, making it hard to identify interactions when such screens are performed in a polyclonal format.

To facilitate the identification of synthetic lethal interactions in mammalian cells with vector-based siRNA libraries, we propose here a technology that we have dubbed 'siRNA barcode screens', which is explained in more detail in BOX 3. Molecular barcodes were first used in yeast genetic screens, in which large numbers of yeast genes were inactivated by insertion of a DNA segment that contained a unique 20-nucleotide sequence (the 'molecular barcode') to mark the individual knockout

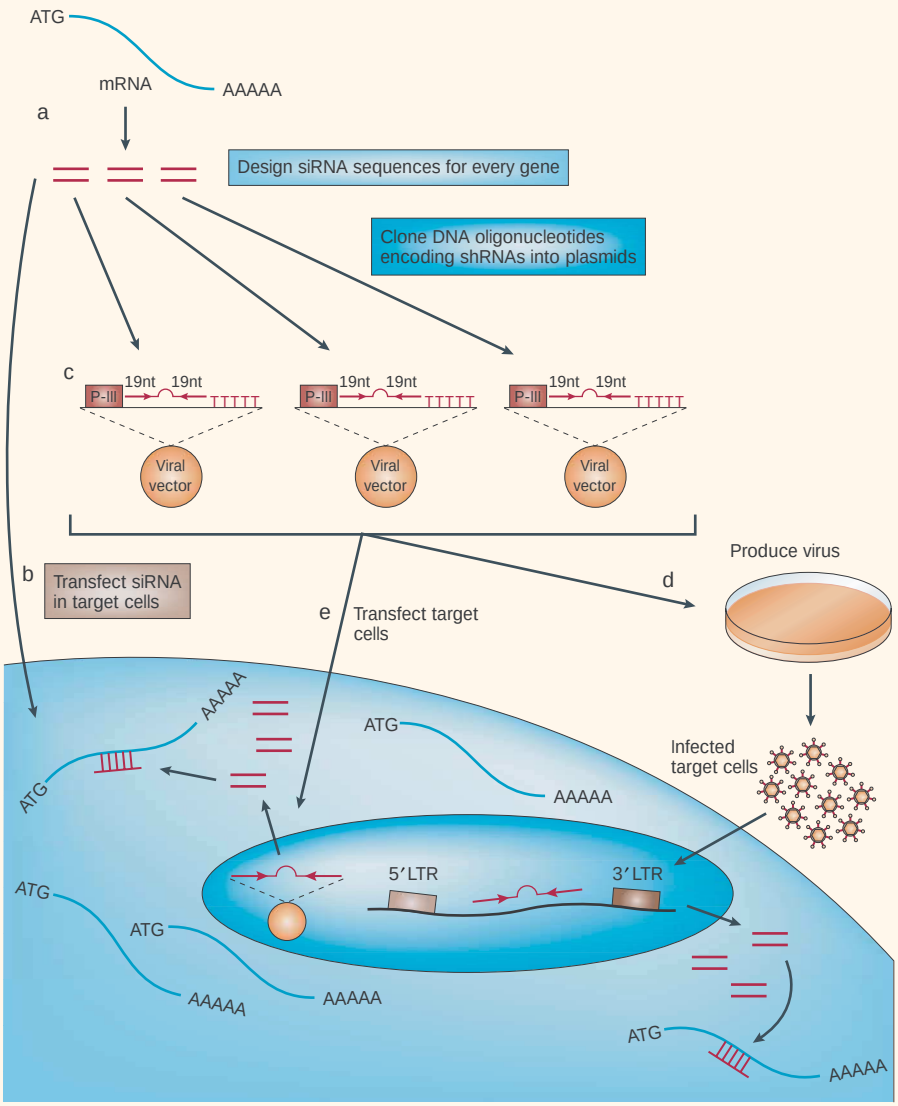
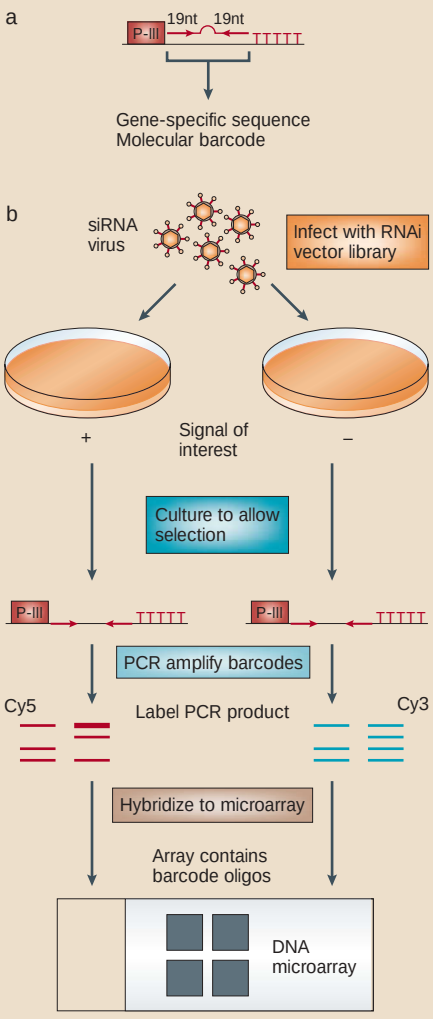


Figure 3 | **Loss-of-function screens using RNAi. a** | To construct a short interfering RNA (siRNA) library, several siRNA sequences (because not every siRNA sequence is active, most groups, including our own, use three or more sequences[51]) are selected to target and inhibit particular genes. **b** | These sequences can be used to generate siRNA molecules that can be chemically synthesized and directly transfected into target cells in a high-throughput manner, and these cells can then be used for phenotypic selection. **c** | Alternatively, DNA oligonucleotides can be designed to mediate expression of short hairpin RNAs (shRNAs) that mimic siRNAs[46–50]. These oligonucleotides are then cloned into modified (retroviral, lentiviral or adenoviral) vectors that contain RNA-polymerase-III (P-III) promoter cassettes. **d** | Retroviral expression of the designed shRNAs results in a population of infected cells in which each infected cell expresses one or more different shRNAs designed against one or more mRNA sequences. **e** | Alternatively, the plasmid vectors can be transfected directly into target cells to drive expression of shRNA molecules in these cells.

yeast strains[65–67]. The molecular barcode serves as a 'strain identifier', which makes it possible to follow the relative abundance of each mutant yeast strain by measuring the abundance of the barcodes in the population. Quantification of the relative abundance of mutant strains under various experimental stress conditions was achieved through polymerase chain reaction (PCR) amplification of the barcodes, followed by hybridization to microarrays containing the barcode sequences.

By analogy to the yeast barcode approach, we noted that the DNA fragment of each siRNA expression vector that encodes the RNA-hairpin transcript has a unique 19-base-pair target-specific sequence. Therefore, introduction of such a vector into mammalian cells results in the creation of a tagged knockdown cell that carries a permanent gene-specific identifier (known as a knockdown identifier; see BOX 3a). This molecular barcode can be recovered by PCR amplification using vector-derived PCR primers that flank the hairpin-encoding DNA sequence. Fluorescent labelling of the PCR product allows it to be identified by hybridization to microarrays that contain the gene-specific knockdown barcode oligonucleotides (see BOX 3a). The relative abundance of each barcoded DNA fragment in the cell population is influenced by the effect that each knockdown vector has on cellular fitness under the experimental conditions and can be quantified using a DNA microarray that contains the barcode DNA fragments. Comparing the effects of knockdown vectors on cellular fitness in pairs of cells that have defined genetic lesions will make it possible to identify synthetic lethal phenotypes in human cells, as siRNA vectors that are specifically lost from one population can be identified in this way.

Box 3 | SiRNA barcode screens



Expression of short hairpin RNA (shRNA) molecules in mammalian cells by stably integrated vectors not only creates a gene-specific knockdown phenotype, but also introduces a gene-specific fingerprint (molecular barcode) in cells that express these siRNAs (see **a**). The 19-base-pair target-gene-specific insert of the siRNA vector is unique in sequence and, therefore, introduction of such a vector into mammalian cells results in the creation of a tagged knockdown cell carrying a permanent gene-specific identifier. This molecular barcode can be recovered by polymerase chain reaction (PCR) amplification using vector-derived PCR primers that flank the hairpin-encoding DNA sequence. When a large collection of siRNA expression vectors is expressed in a population of cells, PCR amplification of the hairpin inserts will result in a mixture of barcode sequences that correspond to the mixture of knockdown vectors that is present in the cell population. The relative abundance of each barcoded siRNA vector in the cell population is influenced by the effect that each knockdown vector has on cellular fitness under the experimental conditions. The abundance of each barcode can be quantified by labelling the PCR product with a fluorescent dye, followed by hybridization to a DNA microarray consisting of barcode-complementary DNA fragments (see **b**) — these oligonucleotides must at least contain the 19 gene-specific nucleotides.

In principle, siRNA barcode screens allow the detection of genetic interactions between large sets of genes and almost any biological signal of interest. A particularly useful application of the siRNA barcode screen is to compare two cell populations, both of which harbour a collection of siRNA vectors, but only one of which is exposed to a biological signal of interest (see **b**) — for example, DNA damage, apoptosis-inducing agents, cytotoxic drugs or inactivation of a tumour-suppressor gene. After PCR amplification, barcoded fragments are labelled with a fluorescent dye (for example, Cy5) and hybridized against barcoded DNA fragments that have been PCR amplified from a control population of cells that contain the same collection of knockdown vectors but that were not exposed to the biological signal of interest. This control population of barcoded DNA sequences is then labelled with a different fluorescent dye (for example, Cy3). Simultaneous hybridization of the Cy5- and Cy3-labelled barcoded DNA fragments allows for the identification of changes in the relative abundance of knockdown vectors in response to the stimulus applied. When a vector is lost specifically in the treated cells, but not in the untreated cells, this indicates that the knockdown vector is synthetically lethal with the stimulus used in the assay. The quantitative nature of DNA-array hybridization makes it possible to analyse large numbers of knockdown vectors in parallel assays, which should greatly facilitate the identification of synthetic lethal interactions in mammalian cell systems.

Future perspectives/conclusions

Many of the cytostatic drugs that are used at present were developed before the molecular basis of cancer was understood in any detail. As a result, most of these drugs are rather non-specific, acting indiscriminately on all rapidly dividing cells. Therefore, the 'therapeutic window' of such drugs (that is, the differential sensitivity of cancer cells versus normal cells) is limited, which causes many of the undesired side effects of anticancer therapy. Over the past decades, we have gained a detailed knowledge of the pathways that are deregulated in cancer and many of the components of these cancer-relevant pathways have been identified. The benefit to the cancer patient of this quantum leap in knowledge has, with some notable exceptions (such as imatinib (Glivec)), been limited. One reason for this is that not all components of cancer-relevant pathways are suitable targets for drug development. Pharmaceutical companies have long known that certain classes of enzymes are more 'druggable' than others. For instance, G-protein-coupled receptors (GPCRs), hydrolases and ion channels are, in general, far more likely to be inhibited by small-molecule compounds than, for instance, transcription factors. So, even though we

have detailed knowledge of a large number of components of cancer-relevant pathways such as the RB, p53 and NF-κB pathways, only a few of the components of these pathways fall into the category of 'druggable' targets. As a result, pharmaceutical companies are still faced with a lack of high-quality drug targets to specifically inhibit those pathways that are deregulated in cancer.

Functional genetic screens are eminently suitable tools to identify novel components of pathways of interest. Indeed, it is probable that many new components of important cancer-relevant pathways can still be identified through functional genetic approaches. A powerful approach in drug-target discovery that is now within reach is to first select a family of 'druggable' genes, generate shRNA vectors against all members of the gene family, and then use such a 'gene family knockdown library' to investigate whether any members of this gene family act in a pathway that is known to be deregulated in cancer. The feasibility of this approach was recently confirmed by showing that one of 50 de-ubiquitylating enzymes acted specifically in the NF-κB pathway[64].

In addition, technologies such as the siRNA barcode screens described here will allow the identification of synthetic lethal interactions, giving rise to unprecedented new classes of potent anticancer drug targets. For example, genetic screens could be used in which the ability of shRNA vectors to modulate cellular fitness in the presence or absence of a second, cancer-specific mutation (for example, mutant *RAS*, overexpressed *MYC*, loss of *TP53*, *PTEN* or *RB*) is compared, allowing the detection of synthetic lethal phenotypes. A similar approach could also be used to identify siRNA vectors that show a genetic interaction with established pharmacological agents, such as anticancer drugs. ShRNA vectors that render cells more sensitive to a pharmacological agent can be readily identified in this way and inhibition of such targets should, in theory, act in synergy with the anticancer drug used in the screen.

We foresee that, together, the advances in functional genetic screening technology in mammalian cells that are described here will have a positive impact on the development of innovative and new classes of anticancer drugs that have more tumour-specific effects and fewer undesired side effects.

Thijn R. Brummelkamp and René Bernards are at the Division of Molecular Carcinogenesis and Center for Biomedical Genetics, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands.
Correspondence to R.B.
e-mail: r.bernards@nki.nl
doi: 10.1038/nrc1191

1. Hanahan, D. & Weinberg, R. A. The hallmarks of cancer. *Cell* **100**, 57–70 (2000).
2. Sherr, C. J. & DePinho, R. A. Cellular senescence: mitotic clock or culture shock? *Cell* **102**, 407–410 (2000).
3. Lin, S. Y. & Elledge, S. J. Multiple tumor suppressor pathways negatively regulate telomerase. *Cell* **113**, 881–889 (2003).
4. Graham, F. L. & van der Eb, A. J. A new technique for the assay of infectivity of human adenovirus 5 DNA. *Virology* **52**, 456–467 (1973).
5. Graham, F. L., van der Eb, A. J. & Heijneker, H. L. Size and location of the transforming region in human adenovirus type 5 DNA. *Nature* **251**, 687–691 (1974).
6. Shih, C. & Weinberg, R. A. Isolation of a transforming sequence from a human bladder carcinoma cell line. *Cell* **29**, 161–169 (1982).
7. Goldfarb, M., Shimizu, K., Perucho, M. & Wigler, M. Isolation and preliminary characterization of a human transforming gene from T24 bladder carcinoma cells. *Nature* **296**, 404–409 (1982).
8. Belt, P. B. *et al.* Efficient cDNA cloning by direct phenotypic correction of a mutant human cell line (HPRT-) using an Epstein–Barr virus-derived cDNA expression vector. *Nucl. Acids Res.* **19**, 4861–4866 (1991).
9. Strathdee, C. A., Gavish, H., Shannon, W. R. & Buchwald, M. Cloning of cDNAs for Fanconi's anaemia by functional complementation. *Nature* **356**, 763–767 (1992).
10. Lo Ten Foe, J. R. *et al.* Expression cloning of a cDNA for the major Fanconi anaemia gene, *FAA*. *Nature Genet.* **14**, 320–323 (1996).
11. de Winter, J. P. *et al.* The Fanconi anaemia group G gene *FANCG* is identical with *XRCC9*. *Nature Genet.* **20**, 281–283 (1998).
12. de Winter, J. P. *et al.* The Fanconi anaemia gene *FANCF* encodes a novel protein with homology to *ROM*. *Nature Genet.* **24**, 15–16 (2000).
13. Pear, W. S., Nolan, G. P., Scott, M. L. & Baltimore, D. Production of high-titer helper-free retroviruses by transient transfection. *Proc. Natl Acad. Sci. USA* **90**, 8392–8396 (1993).
14. Kitamura, T. *et al.* Efficient screening of retroviral cDNA expression libraries. *Proc. Natl Acad. Sci. USA* **92**, 9146–9150 (1995).
15. Zannettino, A. C., Rayner, J. R., Ashman, L. K., Gonda, T. J. & Simmons, P. J. A powerful new technique for isolating genes encoding cell surface antigens using retroviral expression cloning. *J. Immunol.* **156**, 611–620 (1996).
16. Whitehead, I., Kirk, H. & Kay, R. Expression cloning of oncogenes by retroviral transfer of cDNA libraries. *Mol. Cell Biol.* **15**, 704–710 (1995).
17. Shvarts, A. *et al.* A senescence rescue screen identifies BCL6 as an inhibitor of anti-proliferative p19(ARF)-p53 signaling. *Genes Dev.* **16**, 681–686 (2002).
18. Berns, K., Hijmans, E. M., Koh, E., Daley, G. Q. & Bernards, R. A genetic screen to identify genes that rescue the slow growth phenotype of c-*myc* null fibroblasts. *Oncogene* **19**, 3330–3334 (2000).
19. Hitoshi, Y. *et al.* Toso, a cell surface, specific regulator of Fas-induced apoptosis in T cells. *Immunity* **8**, 461–471 (1998).
20. Lewis, P. F. & Emerman, M. Passage through mitosis is required for oncoretroviruses but not for the human immunodeficiency virus. *J. Virol.* **68**, 510–516 (1994).
21. Michiels, F. *et al.* Arrayed adenoviral expression libraries for functional screening. *Nature Biotechnol.* **20**, 1154–1157 (2002).
22. Ziauddin, J. & Sabatini, D. M. Microarrays of cells expressing defined cDNAs. *Nature* **411**, 107–110 (2001).
23. Gudkov, A. V. *et al.* Cloning mammalian genes by expression selection of genetic suppressor elements: association of kinesin with drug resistance and cell immortalization. *Proc. Natl Acad. Sci. USA* **91**, 3744–3748 (1994).
24. Levenson, V. V. *et al.* A combination of genetic suppressor elements produces resistance to drugs inhibiting DNA replication. *Somat. Cell Mol. Genet.* **25**, 9–26 (1999).
25. Novoa, I., Zeng, H., Harding, H. P. & Ron, D. Feedback inhibition of the unfolded protein response by GADD34-mediated dephosphorylation of eIF2α. *J. Cell Biol.* **153**, 1011–1022 (2001).
26. Garkavtsev, I., Kazarov, A., Gudkov, A. & Riabowol, K. Suppression of the novel growth inhibitor p33ING1 promotes neoplastic transformation. *Nature Genet.* **14**, 415–420 (1996).
27. Kimchi, A. Antisense libraries to isolate tumor suppressor genes. *Methods Mol. Biol.* **222**, 399–412 (2003).
28. Deiss, L. P. & Kimchi, A. A genetic tool used to identify thioredoxin as a mediator of a growth inhibitory signal. *Science* **252**, 117–120 (1991).
29. Cohen, O. Kimchi, A. DAP-kinase: from functional gene cloning to establishment of its role in apoptosis and cancer. *Cell Death Differ.* **8**, 6–15 (2001).
30. Deiss, L. P., Feinstein, E., Berissi, H., Cohen, O. & Kimchi, A. Identification of a novel serine/threonine kinase and a novel 15-kD protein as potential mediators of the γ interferon-induced cell death. *Genes Dev.* **9**, 15–30 (1995).
31. Beger, C. *et al.* Identification of Id4 as a regulator of *BRCA1* expression by using a ribozyme-library-based inverse genomics approach. *Proc. Natl Acad. Sci. USA* **98**, 130–135 (2001).
32. Welch, P. J. *et al.* Identification and validation of a gene involved in anchorage-independent cell growth control using a library of randomized hairpin ribozymes. *Genomics* **66**, 274–283 (2000).
33. Kawasaki, H. *et al.* Distinct roles of the co-activators p300 and CBP in retinoic-acid-induced F9-cell differentiation. *Nature* **393**, 284–289 (1998).
34. Colas, P. *et al.* Genetic selection of peptide aptamers that recognize and inhibit cyclin-dependent kinase 2. *Nature* **380**, 548–550 (1996).
35. Xu, X. *et al.* Dominant effector genetics in mammalian cells. *Nature Genet.* **27**, 23–29 (2001).
36. Jorgensen, E. M. & Mango, S. E. The art and design of genetic screens: *Caenorhabditis elegans*. *Nature Rev. Genet.* **3**, 356–369 (2002).
37. Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* **391**, 806–811 (1998).
38. Plasterk, R. H. RNA silencing: the genome's immune system. *Science* **296**, 1263–1265 (2002).
39. Hannon, G. J. RNA interference. *Nature* **418**, 244–251 (2002).
40. Fraser, A. G. *et al.* Functional genomic analysis of *C. elegans* chromosome I by systematic RNA interference. *Nature* **408**, 325–330 (2000).
41. Kamath, R. S. *et al.* Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. *Nature* **421**, 231–237 (2003).
42. Ashrafi, K. *et al.* Genome-wide RNAi analysis of *Caenorhabditis elegans* fat regulatory genes. *Nature* **421**, 268–272 (2003).
43. Pothof, J. *et al.* Identification of genes that protect the *C. elegans* genome against mutations by genome-wide RNAi. *Genes Dev.* **17**, 443–448 (2003).
44. Lum, L. *et al.* Identification of Hedgehog pathway components by RNAi in *Drosophila* cultured cells. *Science* **299**, 2039–2045 (2003).
45. Elbashir, S. M. *et al.* Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* **411**, 494–498 (2001).
46. Brummelkamp, T. R., Bernards, R. & Agami, R. A system for stable expression of short interfering RNAs in mammalian cells. *Science* **296**, 550–553 (2002).
47. Paddison, P. J., Caudy, A. A., Bernstein, E., Hannon, G. J. & Conklin, D. S. Short hairpin RNAs (shRNAs) induce sequence-specific silencing in mammalian cells. *Genes Dev.* **16**, 948–958 (2002).
48. Paul, C. P., Good, P. D., Winer, I. & Engelke, D. R. Effective expression of small interfering RNA in human cells. *Nature Biotechnol.* **20**, 505–508 (2002).
49. Miyagishi, M. & Taira, K. U6 promoter-driven siRNAs with four uridine 3′ overhangs efficiently suppress targeted gene expression in mammalian cells. *Nature Biotechnol.* **20**, 497–500 (2002).
50. Lee, N. S. *et al.* Expression of small interfering RNAs targeted against HIV-1 rev transcripts in human cells. *Nature Biotechnol.* **20**, 500–505 (2002).
51. Paddison, P. J. & Hannon, G. J. RNA interference: the new somatic cell genetics? *Cancer Cell* **2**, 17–23 (2002).
52. Brummelkamp, T. R., Bernards, R. & Agami, R. Stable suppression of tumorigenicity by virus-mediated RNA interference. *Cancer Cell* **2**, 243–247 (2002).
53. Dirac, A. M. & Bernards, R. Reversal of senescence in mouse fibroblasts through lentiviral suppression of p53. *J. Biol. Chem.* **278**, 11731–11734 (2003).
54. Abbas-Terki, T., Blanco-Bose, W., Deglon, N., Pralong, W. & Aebischer, P. Lentiviral-mediated RNA interference. *Hum. Gene Ther.* **13**, 2197–2201 (2002).
55. Rubinson, D. A. *et al.* A lentivirus-based system to functionally silence genes in primary mammalian cells, stem cells and transgenic mice by RNA interference. *Nature Genet.* **33**, 401–406 (2003).
56. Scherr, M., Battmer, K., Ganser, A. & Eder, M. Modulation of gene expression by lentiviral-mediated delivery of small interfering RNA. *Cell Cycle* **2**, 251–257 (2003).
57. Qin, X. F., An, D. S., Chen, I. S. & Baltimore, D. Inhibiting HIV-1 infection in human T cells by lentiviral-mediated delivery of small interfering RNA against *CCR5*. *Proc. Natl Acad. Sci. USA* **100**, 183–188 (2003).
58. McCaffrey, A. P. *et al.* RNA interference in adult mice. *Nature* **418**, 38–39 (2002).
59. Kunath, T. *et al.* Transgenic RNA interference in ES cell-derived embryos recapitulates a genetic null phenotype. *Nature Biotechnol.* **21**, 559–561 (2003).
60. Jackson, A. L. *et al.* Expression profiling reveals off-target gene regulation by RNAi. *Nature Biotechnol.* **21**, 635–637 (2003).

61. Semizarov, D. *et al.* Specificity of short interfering RNA determined through gene expression signatures. *Proc. Natl Acad. Sci. USA* **100**, 6347–6352 (2003).
62. Chi, J. T. *et al.* Genomewide view of gene silencing by small interfering RNAs. *Proc. Natl Acad. Sci. USA* **100**, 6343–6346 (2003).
63. Bridge, A. J., Pebernard, S., Ducraux, A., Nicoulaz, A. L. & Iggo, R. Induction of an interferon response by RNAi vectors in mammalian cells. *Nature Genet.* **34**, 263–264 (2003).
64. Brummelkamp, T. R., Nijman, S. M. B., Dirac, A. M. G. & Bernards, R. Loss of the cylindromatosis tumour suppressor inhibits apoptosis by activating NF-κB. *Nature* **424**, 797–801 (2003).
65. Shoemaker, D. D., Lashkari, D. A., Morris, D., Mittmann, M. & Davis, R. W. Quantitative phenotypic analysis of yeast deletion mutants using a highly parallel molecular bar-coding strategy. *Nature Genet.* **14**, 450–456 (1996).
66. Winzeler, E. A. *et al.* Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* **285**, 901–906 (1999).
67. Giaever, G. *et al.* Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* **418**, 387–391 (2002).
68. Hannon, G. J. *et al.* MaRX: an approach to genetics in mammalian cells. *Science* **283**, 1129–1130 (1999).
69. Koh, E. Y., Chen, T. & Daley, G. Q. Novel retroviral vectors to facilitate expression screens in mammalian cells. *Nucl. Acids Res.* **30**, e142 (2002).
70. Hua, X., Liu, X., Ansari, D. O. & Lodish, H. F. Synergistic cooperation of TFE3 and smad proteins in TGF-β-induced transcription of the plasminogen activator inhibitor-1 gene. *Genes Dev.* **12**, 3084–3095 (1998).
71. Hudson, J. D. *et al.* A proinflammatory cytokine inhibits p53 tumor suppressor activity. *J. Exp. Med.* **190**, 1375–1382 (1999).
72. Maestro, R. *et al.* Twist is a potential oncogene that inhibits apoptosis. *Genes Dev.* **13**, 2207–2217 (1999).
73. Yamaoka, S. *et al.* Complementation cloning of NEMO, a component of the IκB kinase complex essential for NF-κB activation. *Cell* **93**, 1231–1240 (1998).

Acknowledgements
We thank R. Beijersbergen for helpful discussions and critical reading of this manuscript and apologize to our colleagues for omission of relevant work due to space constraints. The work of the authors was supported by grants from the Dutch Cancer Society, The Netherlands Organization for Scientific Research and the Center for Biomedical Genetics.

Online links

DATABASES
The following terms in this article are linked online to:
LocusLink: http://www.ncbi.nlm.nih.gov/LocusLink/
ARF | *β-catenin* | *BRCA1* | CBP | *CYLD* | *env* | *gag* | *Hedgehog* | ID4 | ING1 | *KRAS* | MDM2 | *MYC* | NF-κB | p300 | p53 | *pol* | *PPAN* | *PTEN* | *RAF1* | RB

OPINION

Proteomic analysis of cancer-cell mitochondria

Mukesh Verma, Jacob Kagan, David Sidransky and Sudhir Srivastava

Mitochondrial dysfunction and mutations in mitochondrial DNA have been frequently reported in cancer cells. Mitochondrial gene-expression signatures of transformed cells have been identified; however, the phenotypic effects of these genetic alterations remain to be established. Identification of mitochondrial proteins that are aberrantly expressed in cancer cells has been made possible by the recent development of mitochondrial functional proteomics and could identify new markers for early detection and risk assessment, as well as targets for therapeutic intervention.

Each human cell contains several hundred copies of mitochondrial DNA (mtDNA), which encodes 13 Kreb's cycle and respiratory-chain subunits, 22 transfer RNAs and 2 ribosomal RNAs (rRNAs) — 12S and 16S[1]. The main function of mitochondria is to produce cellular energy and, during this process, acetyl-coenyme A — the breakdown product of fats and sugars — is passed through the Kreb's cycle to generate electrons. These, in turn, are transferred to the proteins of the respiratory chain in the inner mitochondrial membrane. This generates a proton gradient between the mitochondrial matrix and the intermembrane space, which is used to generate ATP. In addition to serving as the main intracellular source of energy of the cell, mitochondria regulate several cellular processes that are linked to apoptosis, which include electron transport and energy metabolism. They are also the storage site for a number of soluble proteins that mediate apoptosis, including cytochrome *c*, certain procaspases and apoptosis-inducing factor. Mitochondria can activate apoptosis by releasing these factors from their intermembrane space into the cytoplasm, and also by altering the cellular redox potential[2–4]. During this process, reactive oxygen species (ROS) are generated that can serve as crucial pro-apoptotic factors. In fact, recent studies have indicated that mitochondria can undergo self-apoptosis — termed 'mitoptosis'[5]. Therefore, mitochondria not only serve as a powerhouse of energy, but also as a supplier of apoptosis-inducing signals (FIG. 1).

We have recently gained much insight into the connection between mitochondrial dysfunction, deregulation of apoptosis and tumorigenesis. Further information about mitochondrial proteins that are involved in cancer progression could lead to new markers for early cancer detection and prevention. Recent advances in mitochondrial genomic and proteomic techniques have greatly improved our ability to identify these types of markers. Here, we discuss the current status of mitochondrial proteomics and associated challenges in cancer detection.

MtDNA mutations and cancer

Some commonly used DNA-based markers of cancer include loss of heterozygosity (LOH), microsatellite instability (MSI), DNA hypermethylation, nuclear and mtDNA mutations, and detection of cancer-associated viral DNA. Mitochondria contain their own DNA, which is replicated and transcribed semi-autonomously. MtDNA is particularly susceptible to damage by environmental carcinogens because it contains no introns, has no protective histones or non-histone proteins, and is exposed continuously to endogenous ROS. In fact, the frequency of mtDNA mutations in cancer cells has been reported to be tenfold higher than that of nuclear DNA mutations[6,7].

The mitochondrial D loop is a section of the mitochondrial genome that is thought to be involved in replication and that contains short poly-pyrimidine tracts. The D loop seems to be a hot spot for mutations, although mutations have been identified throughout the mitochondrial genome[1]. Many of these mutations seem to be shared by most or all of the mitochondria that are contained in a cancer cell — that is, the mutations are at homoplasmic levels — providing a potential advantage for molecular detection. Many cancer-associated mutations have also been identified in the 12S and 16S rRNA genes[8,9,10].

A number of mutations, deletions and insertions in the mitochondrial genome have been associated with specific cancers. Microsatellite DNA alterations in the non-coding D loop and insertions and deletions have been observed in the mtDNA genome, particularly in colorectal and gastric cancer. In addition, frameshift mutations have been identified in the D loop of pre-neoplastic head and neck cancer cells[11]. The D310 D-loop variants harbour large C-tract deletions that are likely to interfere with the initiation of mtDNA replication[12] and mitochondria that undergo rapid replication in chronic hepatitis can acquire and accumulate DNA damage and mutation more readily than those maintained under resting conditions. Base substitutions in the *MTND1*, *MTND4*, *MTND5* and *cytochrome b* genes, or their noncoding region, can lead to breast cancer[13]. MtDNA mutations in genes that