Towards systematic functional characterization of cancer genomes

Jesse S. Boehm and William C. Hahn*‡§*

Abstract | Whole-genome approaches to identify genetic and epigenetic alterations in cancer genomes have begun to provide new insights into the range of molecular events that occurs in human tumours. Although in some cases this knowledge immediately illuminates a path towards diagnostic or therapeutic implementation, the bewildering lists of mutations in each tumour make it clear that systematic functional approaches are also necessary to obtain a comprehensive molecular understanding of cancer. Here we review the current range of methods, assays and approaches for genome-scale interrogation of gene function in cancer. We also discuss the integration of functional-genomics approaches with the outputs from cancer genome sequencing efforts.

Oncogenes
Genes that are somatically mutated or amplified in tumours, are required for the survival of tumours that harbour the oncogene and cause transformation in a cell or animal model.

Tumour suppressor genes
Genes that show loss of heterozygosity in tumours and usually regulate cell survival.

**Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, Massachusetts 02142, USA.*
‡Department of Medical Oncology, Center for Cancer Genome Discovery, Dana-Farber Cancer Institute, 450 Brookline Avenue, Dana 1538, Boston, Massachusetts 02115, USA.
§Departments of Medicine, Brigham and Women's Hospital and Harvard Medical School, 75 Francis Street, Boston, Massachusetts 02115, USA.
Correspondence to W.C.H.
e-mail: william_hahn@dfci.harvard.edu
doi:10.1038/nrg3013

The recent application of whole-genome tools to characterize normal and cancer genomes now provides the means to enumerate the multiple somatic genetic and epigenetic alterations that occur in cancer[1]. Indeed, it is likely that the interrogation of somatic alterations will eventually be deployed in the characterization of many, if not all, cancers and will increasingly influence diagnostic and therapeutic decisions. Such efforts have already facilitated the discovery of novel oncogenes and tumour suppressor genes and, in some cases, have led to the identification of cancer subtypes or prediction of the clinical response based on molecular, rather than anatomic, criteria[2,3].

However, as most cancers exhibit evidence of genomic instability, only a subset of the large number of somatic alterations harboured by each tumour contributes to cancer phenotypes. Although recurrent mutations that occur in a statistically significant subset of cancers are likely to have a direct role in cancer development, even low-frequency mutations may be both biologically and clinically important. This is demonstrated by the finding that the small percentage of non-small-cell lung cancers that harbour translocations of anaplastic lymphoma receptor tyrosine kinase (*ALK*) respond dramatically to small-molecule inhibitors of ALK[4,5]. Therefore, identifying the subset of these mutant genes that are involved in cancer initiation and/or progression requires a complementary approach to define the genes that have key roles in cancers.

Functional studies can provide initial clues to the mechanistic role of mutations in cancer initiation and/or progression; this information will inform our understanding of the biological basis of cancer and direct our efforts to develop therapeutic strategies. Moreover, as oncogenes and tumour suppressor genes operate within pathways and networks, functional studies provide the means to decipher the signalling pathways affected by somatically altered genes, as well as the means to understand the differences between the normal and dysregulated function (or functions) of these networks in cancer. As such, functional studies in cancer have the potential to uncover other components of these pathways that have essential roles and that may serve as potential therapeutic targets but are not necessarily mutated in cancer genomes.

Studies of specific genes in particular cancer lineages, tumour contexts or cancer phenotypes have provided much of our current knowledge of cancer pathogenesis. Indeed, focused functional inquiries into experimental models that recapitulate tumour biology will continue to be an important strategy to study the molecular basis of cancer. However, the pace of discovery of somatically altered genes, pathways and networks in cancer requires the deployment of efficient, systematic methods to interrogate gene function. Systematic approaches to characterize the function of genes will catalogue candidate genes that are involved in producing cancer-related phenotypes (FIG. 1, left) and will define the relationship of specific genes with known cancer-related molecules or signalling pathways — for example, by creating signatures of their perturbation in model systems (FIG. 1, right).

The development of genome-scale molecular tools to perturb gene function in mammalian cells, as well as assays with appropriate sensitivity, specificity,

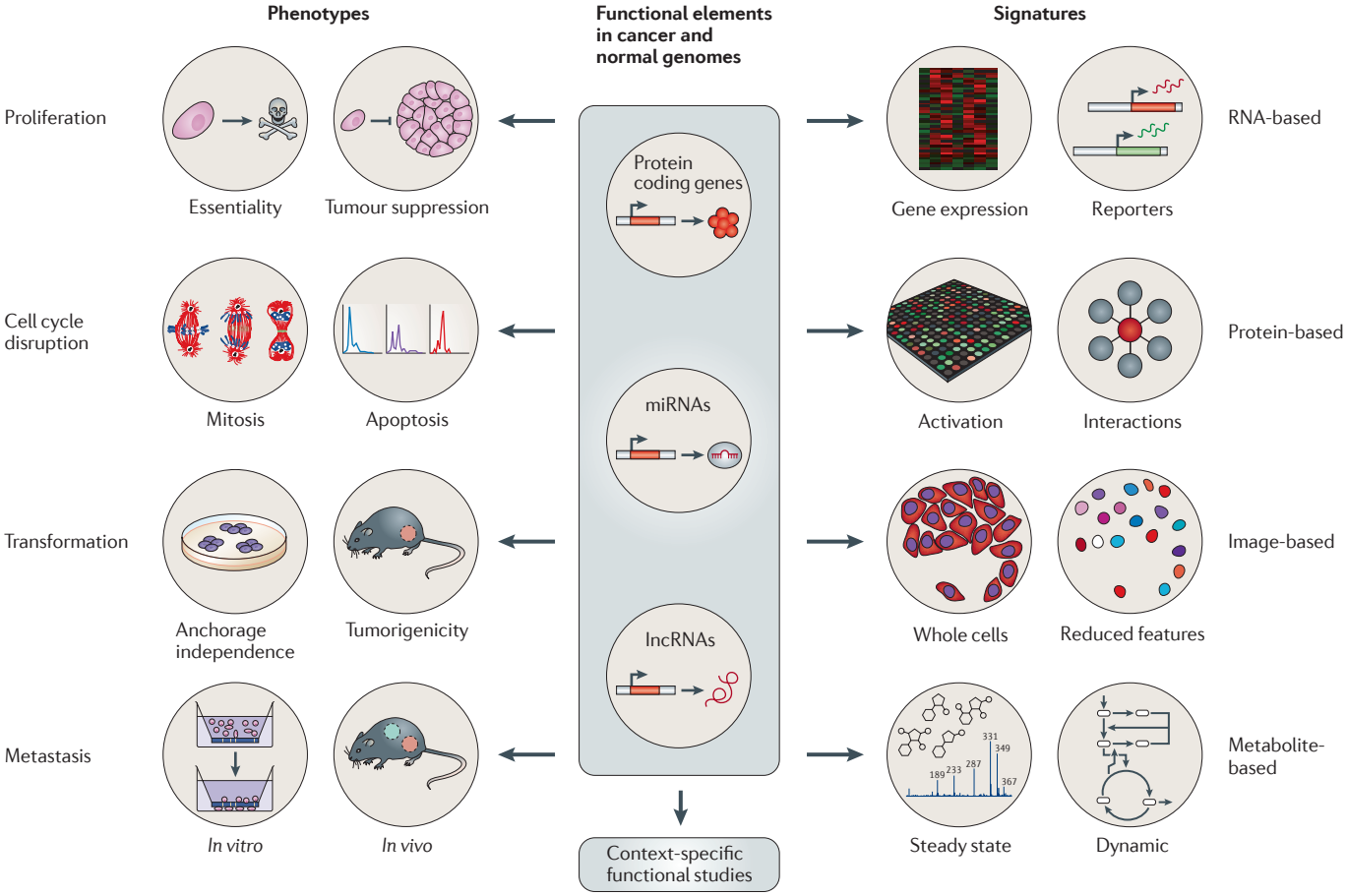


Figure 1 | **Overview of cancer functional genomics.** A schematic representation of commonly used approaches to characterize the function of elements that are encoded in cancer and normal genomes. As efforts associated with systematically identifying such elements proceed, the identities and sequences of all protein-coding genes and their isoforms, microRNAs (miRNAs) and long non-coding RNAs (lncRNAs), as well as other elements (not pictured), are rapidly being discovered. One method of functionally characterizing these elements is by producing molecular signatures of their perturbation and connecting these signatures to those of known cancer pathways, molecules or phenotypes. Some examples of categories of signature-based approaches to guide functional gene annotation in cancer are displayed on the right. Genes may also be functionally annotated by their contribution to cancer-associated phenotypes (examples are shown on the left). High-throughput phenotype- and signature-based approaches to functional cancer gene annotation are complemented by deeper investigation in certain tissue-specific contexts and *in vivo* models.

throughput and cost, is necessary to enable large-scale systematic functional studies. Near-genome-scale functional interrogation in mammalian cells is now possible with technical advances in both areas, including improved open reading frame (ORF), RNA interference (RNAi) and chemical libraries, in addition to more sophisticated methods for identifying phenotypic changes. In this Review, we provide an overview of progress in developing and implementing methods to interrogate gene function at increased scale, demonstrate the synergy possible from combining structural and functional cancer genome efforts and discuss emerging areas of investigation.

Tools for systematic functional analyses

The basic tenet of functional-genomics studies is that, by perturbing the expression or inhibiting the function of a gene product in an assay, one can gain insight into its biological function (or functions). In this section, we introduce tools for altering expression or function at genome scale. An overview of these tools, and the experimental formats in which they may be deployed, is depicted in FIG. 2.

Systematic mutagenesis. Transposons and retroviruses can now be used to disrupt gene function at genome scale in mice. Although transposons are generally inactive in mammalian cells, several groups have engineered activated, recombinant transposons that, when introduced into the murine genome, permit the generation of mice in which genes are activated or inactivated owing to integration events in, or near, genes[6]. Using either the fish-derived TC1/mariner transposon (*Sleeping Beauty*)[7] or the cabbage-looper moth *Trichoplusia ni* transposon (*piggyBac*)[8–10], these investigators have generated large cohorts of mice that develop cancers in many different

Open reading frame (ORF). The coding sequence of a transcript without 5′ or 3′ sequences.

RNA interference (RNAi). The process by which endogenous or exogenous dsRNA molecules lead to interference with gene expression.

Functional-genomics studies. The manipulation of gene expression or function at large scale, usually using high-throughput approaches.

Transposons. DNA elements that can move to new positions within the genome of a single cell.

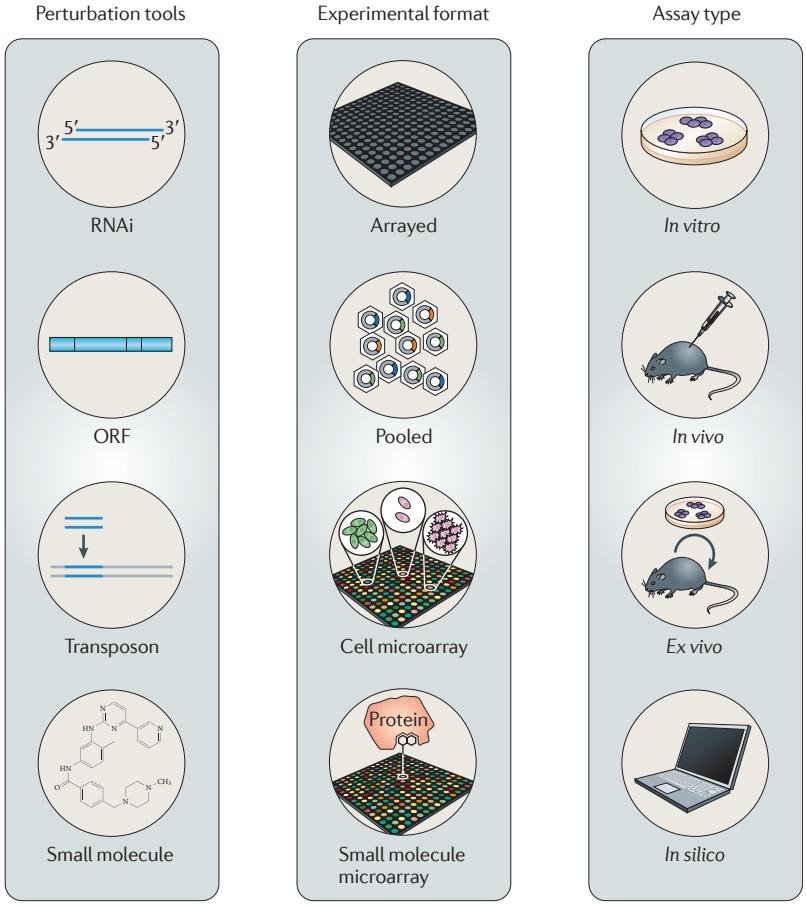


Figure 2 | **Tools and formats of cancer functional-genomics experiments.** High-throughput cancer functional-genomics experiments require the selection of genetic or chemical tools for perturbation (left) as well as an experimental format (centre) and assay type (right). Designing screens involves combining tools, formats and assay types, together with a suitable readout for the biological process of interest. ORF, open reading frame; RNAi, RNA interference.

organs over time. For example, activation of the *Sleeping Beauty* transposon in the gastrointestinal tract led to the development of the full range of hyperplastic to malignant lesions caused by the inactivation of known tumour suppressor genes, such as *Apc*, *Pten* and *Smad4* (REF. 7). In addition, activation of the *Sleeping Beauty* transposon identified new candidate genes in colon cancer pathogenesis, including *Poli*, *Ptprk* and *Rspo2*. Similarly, activation of the *piggyBac* transposon led to the identification of *Spic* and histone deacetylase 7 (*Hdac7*) as genes involved in haematopoietic tumours[9]. The generation of mice in which these transposons are conditionally activated in different genetic contexts and tissues will provide the means to identify novel genes involved in tumour initiation and maintenance.

Retroviruses have also been used for large-scale insertional mutagenesis in mice[11]. For example, the introduction of retroviruses into mice lacking either of the tumour suppressors alternative reading frame 1 (ARF1; encoded by *Cdkn2a*) or p53 has led to the identification of a number of candidate oncogenes and tumour suppressor genes[12]. Because retroviruses integrate at different sites from transposons, retrovirus- and transposon-based approaches are complementary. Retroviruses have also recently been used to perform a loss-of-function screen in a chronic myeloid leukaemia cell line that is haploid at all chromosomes except for chromosome 8 (REF. 13). This study focused on the identification of host factors that are necessary for viral infections, but it could be used to study many other phenotypes, including those relevant to cancer[13]. Although this approach is currently limited to cell lines that are haploid for specific chromosomes, this work provides a powerful experimental system in which to create null alleles in human cells.

RNA interference. In mammalian cells, small dsRNAs, such as short interfering RNAs (siRNAs) and microRNAs (miRNAs), regulate gene expression through specialized gene-silencing enzymatic complexes. This machinery can be harnessed to target individual genes or to target large numbers of genes in screens. The specificity of RNAi depends on sequence homology, which permits the production of prospectively designed reagents. Indeed, academic and commercial groups have produced synthetic small dsRNA molecules and plasmid-encoded short hairpin RNAs (shRNAs) that target most of the predicted human or murine genes. Compared to insertional mutagenesis, RNAi offers the advantage of speed, flexibility and convenience; however, RNAi generally induces gene suppression and thus experiments using RNAi are more similar to the use of hypomorphic alleles than null alleles.

Synthetic siRNAs targeting individual genes, gene sets or entire transcriptomes are available for purchase from multiple vendors and can be introduced into mammalian cells by several means of gene transfer, including lipid-based transfection reagents and electroporation. Typically, siRNAs induce substantial gene suppression for several days and are widely used to assess short-term phenotypes in cells that can be readily transfected. Genome-scale collections of siRNAs have facilitated large arrayed screens (discussed below) to identify genes involved in phenotypes such as apoptosis[14] and synergy with chemotherapy reagents[15]. The key considerations in these types of experiments are the robustness of the assay used to measure a specific phenotype or pathway activity and whether the RNAi reagents can be introduced efficiently into the target cells.

Vector-based methods to deliver shRNAs permit the stable expression of RNAs to enable longer term experiments. The use of different vectors, such as retroviruses, can expand the range of types of cells that can be studied, including primary cells, non-dividing cells and cells within an organism. Near-genome-scale collections are now available in several types of expression vectors. The most commonly used expression systems are: retrovirus-based systems in which the expression of shRNAs is driven by RNA polymerase (Pol) III promoters[16,17]; systems driven by Pol II promoters in which the shRNA is placed in the context of and is processed by the microRNA machinery[18]; or systems in which antiparallel RNAs are

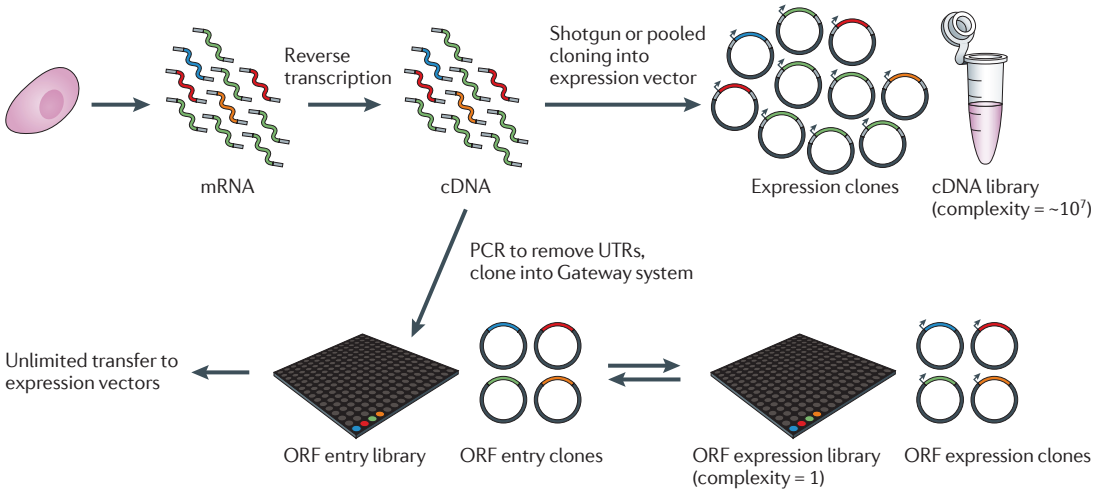
Short interfering RNAs
(siRNAs). RNA molecules that are capable of inducing RNA interference.

MicroRNAs
(miRNAs). MicroRNAs are short RNA molecules that regulate gene expression through gene silencing and translational repression.

Short hairpin RNAs
(shRNAs). An RNA interference-inducing molecule that folds back onto itself to create a hairpin structure.

Arrayed screens
Functional-genomics screens in which perturbations are individually performed.

Box 1 | **cDNA and ORF libraries**



Libraries of cDNA or purified open reading frames (ORFs) derived from such cDNAs and encoded in expression vectors are powerful tools for overexpression experiments. To make these resources, a population of mRNA molecules (see figure) is isolated from cells (primary or cancer) or from tissue and a reverse-transcription reaction produces cDNA molecules, including UTRs (grey sections of molecules). A cDNA library can rapidly be created from this material by shotgun or pooled cloning into expression vectors. Pooled cDNA libraries have a high complexity (typically a mixture of ~10^7 molecules) and, therefore, only cancer functional-genomics experiments with a sensitivity equal to or greater than the library complexity may be used. Clone representation is proportional to cellular expression and is dependent both on the size and GC content of the gene. Thus, gene representation in the pool is unequal. Lastly, UTRs flank clones in cDNA libraries, precluding some applications. By contrast, ORF libraries are typically produced by removing UTRs and cloning into a flexible clone-shuttling system, such as Gateway (Invitrogen), and they allow for greater experimental control of gene expression. Clones are typically produced in arrayed format and thus have equal representation and low complexity. Gateway Entry libraries can be used for flexible clone transfer into unlimited numbers of expression vectors. As clones lack UTRs, expression libraries can easily incorporate epitope or functional tags to enable downstream applications.

simultaneously expressed from Pol III promoters and hybridize to form dsRNAs[19]. In addition, the use of promoters that permit inducible regulation of shRNA expression allows the regulation of gene expression both *in vitro* and *in vivo*[20].

Despite the capacity to prospectively design RNAi reagents, not all constructs in available libraries are equally effective at eliciting gene suppression, owing to an incomplete understanding of how small-RNA-mediated silencing occurs *in vivo*. For this reason, most collections are composed of multiple independent shRNAs (or siRNAs) targeting each gene to maximize the likelihood that more than one effective shRNA is included. Although recent advances in reagent design are leading to improved RNAi tools[21], it is not yet possible to carry out saturating genetic screens in mammalian models. In addition, off-target effects of RNAi — which are not fully understood[22] — occur and make it essential to use multiple controls and methods to validate findings from primary screens.

cDNA, ORF and miRNA expression libraries. Overexpression systems have long been used to explore gene function. For example, fragmented genomes or cDNA-expression libraries produced from mRNA from specific cells or tissues have been used to identify many receptors and ligands[23], oncogenes[5,24] and growth factors[25]. Despite these successes, the use of libraries derived from the pooled reverse transcription of mRNA is limited by the unequal representation of genes as a result of differences in expression in the donor cell or tissue or representation after reverse transcription (BOX 1). The Mammalian Gene Collection (MGC) project has created a large collection of sequence-confirmed arrayed cDNAs that now serve as useful templates for downstream applications[26].

Over the past decade, increasingly complete collections of ORFs of human genes have been created[27–30], including a recent collection of over 16,000 fully sequenced human ORFs. This represents over 13,500 human genes in a high-titre, lentiviral expression vector system enabling high-throughput ORF screens in mammalian cells[31,32] (BOX 1). However, the largest currently available ORF libraries are not yet truly comprehensive, as a substantial fraction of genes and isoforms have not yet been isolated and a fraction of the clones harbour nucleotide alterations.

Gain-of-function approaches can also be used to study the roles of non-coding RNAs such as miRNAs in cancer. Approximately 500 human miRNAs have been described to date, although the targets of most of these miRNAs remain undefined[33] and individual miRNAs can coordinately regulate the expression of many genes. MicroRNAs contribute to malignant transformation as both oncogenes and tumour suppressor genes[34]. Expression libraries that are composed of miRNAs have now been made and are used in phenotypic screens[35]. For example, a screen showed that miR-372 and miR-373

Off-target effects
A term that refers to a phenotype that is not related to perturbation of the intended target of a short interfering RNA (siRNA) or small molecule.

Transformation
The process by which a normal cell acquires cellular phenotypes of a cancer cell.

permit cell proliferation in the setting of a strong inducer of proliferative arrest[35]. The relatively small number of miRNAs (compared to protein-coding genes) facilitates the use of these libraries in both *in vitro* and *in vivo* screens.

Chemistry-based approaches. Screens involving small molecules are another effective approach for identifying novel gene products and pathways involved in cancer. Screening small molecules in biochemical- or cell-based assays has long been used to identify potential chemotherapeutic agents; however, advances in biochemical methods for identifying binding partners, innovative new screening paradigms and novel libraries of chemicals now permit the use of these approaches to study cancer phenotypes.

One option for using small molecules for cancer functional genomics involves screens with compounds that have known mechanisms of action and/or targets, so that experimental results can be directly traced to gene function. For example, screening a panel of kinase inhibitors across a panel of 500 cell lines led to the identification of kinases required for the proliferation of specific cell lines[4,36,37]. Deploying other such collections of small-molecule probes with a known function and a known target (or targets) in high-throughput screens across hundreds of cell lines is an emerging area of active investigation[38].

A complementary approach involves the creation and deployment in assays of small-molecule collections that cover an increasingly comprehensive chemical space, followed by the identification of proteins mediating the function of top-scoring candidates. Several academic groups have created small-molecule collections that are composed of molecules with novel chemical properties, including increased chemical complexity, chirality and easily modified side chains. This not only expands the types of molecules that can be screened but also facilitates the identification of protein targets through the addition of capture tags[39]. By combining these reagents with well-characterized cell lines, several laboratories have identified novel mechanisms that regulate signalling. For example, using innovative chemical libraries and small-molecule microarrays, small molecules that inhibit Hedgehog and Notch signalling have been identified[40,41], and these molecules have already proven useful in dissecting the function of these pathways in cancer.

In addition to facilitating the discovery of novel small molecules that influence specific pathways, several groups have devised innovative assays to identify small molecules that perturb complex biological phenotypes. For example, an image-based screen to identify regulators of stem cell function found the small molecule stauprimide, which interacts with and inhibits nuclear localization of the nucleoside diphosphate kinase B (NDKB), which, in turn, affects MYC expression[42]. The combination of innovative chemical libraries, together with increasingly informative biological assays, provides investigators with the means to not only gain insight into the mechanisms responsible for specific phenotypes but also to identify chemical tools that permit further investigation and that may be the basis for lead compounds.

Assay design and strategy

The tools described in the previous section can be used in a wide range of formats, assays and contexts. Ideally, assay selection is dictated by the biological phenotype under study; however, technical details and limitations also influence experimental design. The key considerations in designing screens are whether the assay truly recapitulates the phenotype of interest and whether the assay is sufficiently robust to detect differences under the experimental conditions.

Experimental format. Two general formats have been used for genetic screens in mammalian cells: arrayed and pooled formats (FIG. 2). Both can be used for high-throughput screening and have distinct advantages and limitations.

Arrayed-format experiments involve screening an array of spatially segregated wells or spots that each contain a single reagent (such as an siRNA, ORF or chemical). These experiments use miniature phenotypic or pathway-specific readouts (FIG. 1). Arrayed screens provide the advantage of identifying subtle phenotypes because the effect of each reagent is interrogated separately and can be coupled with a large number of complex phenotypic assays[17,43,44]. Cell microarrays are specialized forms of arrayed-format screens in which viral reagents are attached to a solid surface and cells are plated over the entire surface[45]. These types of arrayed screens permit the screening of large numbers of reagents, as one is not limited by the number of wells in a plastic plate. However, owing to the small number of cells that contact each reagent, cell microarrays are most effective when coupled with a detector or image-based assay[45]. A complementary approach involves spotting small molecules in an array format, incubating with the purified protein of interest and identifying binding compounds. These small-molecule microarrays constitute an efficient technique that holds considerable promise for the identification of modulators of many protein types, including transcription factors[46].

Although arrayed-format screens are powerful, the cost and complexity of performing large screens limits their utility. However, vector-based shRNA and ORF libraries can also be used in pooled formats that permit the parallel interrogation of large numbers of genes in a single experiment. In these types of experiments, pools of retroviruses encoding shRNAs or ORFs are introduced into target cells. The infected cells are then subjected to a selective pressure, such as treatment with a drug or passage in culture, and the abundance of each shRNA is quantified before and after selection by using microarrays or deep sequencing.

Such screens can use positive selection or negative selection approaches. Positive selection screens examine enrichment of integrated shRNAs or ORFs in cells under selective pressure, such as drug treatment[47]. When performing such screens in small numbers of cancer

cell lines, one must eliminate the possibility that genetic alterations that are harboured by the cell line cooperate with the introduced reagent to drive the observed phenotype.

Technical advances that allow the measurement of shRNAs that are depleted in pooled screens mean that negative selection screens can now be performed. For example, a pooled shRNA screen in U2OS cells (an osteosarcoma cell line) identified genes involved in the response to ionizing radiation[48] and intrastrand crosslink repair[49]. Additionally, several groups have reported successes in using genome-scale pooled screens to identify essential and therefore negatively selected genes[17,18,50–55]. Although powerful, detecting the depletion of specific shRNAs in negative selection screens requires careful optimization with positive controls to verify that one has defined conditions that have appropriate signal-to-noise properties.

For non-mammalian model organisms, loss-of-function screens are often performed by comparing the same cell under different experimental conditions or by comparing cells that differ by a single allele. However, the considerable heterogeneity that is inherent in mammalian systems limits the power of primary screens performed in paired cell lines to unequivocally identify genes whose function distinguishes the two classes (BOX 2). For example, cancer cell lines harbour many genetic alterations, even when they have been engineered to express a specific allele, and normal human cells — particularly epithelial cells — are difficult to propagate under the necessary conditions for performing large-scale screens. Moreover, the large number of simultaneous comparisons requires that one must account for multiple-hypothesis testing when interpreting the outcomes of such genome-scale experiments. One approach to address each of these issues is to perform secondary experiments to find candidates that give similar phenotypes in several cell lines or contexts. Alternatively, high-throughput, pooled-format screening experiments on large panels of well-characterized cell lines provide increased power to identify true positives (BOX 2).

In addition to providing the means to perform genome-scale screens, pooled formats can be used *in vivo*. Several investigators have developed targeted shRNA libraries that are composed of up to 1,000 shRNAs; the shRNAs are introduced into freshly isolated cells *in vitro* and then the cells are transplanted into recipient animals to identify genes that are involved in survival or tumorigenesis *in vivo*. For example, using pluripotent cells from the liver[56] or haematopoetic system[57,58], investigators have identified candidate tumour suppressor genes such as *Sfrp1*, *Numb*, *Mek1* (also known as *Map2k1*) and angiopoietin 2 (*Angpt2*)[57], as well as several genes that regulate actin expression and are required for lymphoma survival[58]. One technical limitation of these experiments is that the number of shRNAs that are assayed *in vivo* in any one animal is limited, owing in part to the lower sensitivity of many *in vivo* readouts. However, coupled with recent advances in making vectors that allow for inducible expression[59], these approaches facilitate experiments to study the effects of gene suppression in tissue- and time-specific contexts.

Assays. The most commonly measured cancer phenotypes are proliferation and survival, and these phenotypes have been used during the development of methods and reagents for proof-of-principle experiments. For example, shRNA libraries have been deployed in arrayed-format screens to investigate the genes involved in producing many cancer phenotypes, such as dysregulation of cell division[60], abnormal proteosome function[50] and aniokis[61]. The development of new imaging analysis tools[62] will facilitate the development of assays that measure the activity of specific molecules as well as complex phenotypes (BOX 3).

Cell lines that depend on particular pathways for survival enable assays to identify genes involved in cancer. For example, arrayed-format screens have identified genes that regulate the p53 pathway[63] or the phosphatidylinositol 3-kinase (PI3K) pathway[64], as well as regulators of RAS signalling[65]. An alternative approach is to use reporter constructs to measure the activity of a specific signalling pathway. For example, a reporter construct that measured the transcriptional activity of the β-catenin–T cell factor 4 (TCF4) complex has been used to identify new regulators of the β-catenin pathway[66].

Robust assays are necessary in order to overcome the experimental variability that accompanies large-scale screens. For this reason, many investigators have used positive selection screens, such as the induction of anchorage-independent growth, to identify genes involved in cell transformation. For example, using a retrovirally delivered cDNA library, *DRIL1* (also known as *ARID3A*) was identified as a gene that rescued senescence induced by the expression of oncogenic RAS[67]. Using RNAi, *PITX1* was found as a gene whose suppression induces RAS activation through the RASAL1 protein in colon cancer cell lines[65]; additionally, a network of epigenetically regulated genes was identified as necessary for RAS-mediated epigenetic silencing of the pro-apoptotic gene *FAS*[68]. Similarly, cell-separation assays coupled with RNAi screening have been used to identify genes whose suppression increases invasion[69,70].

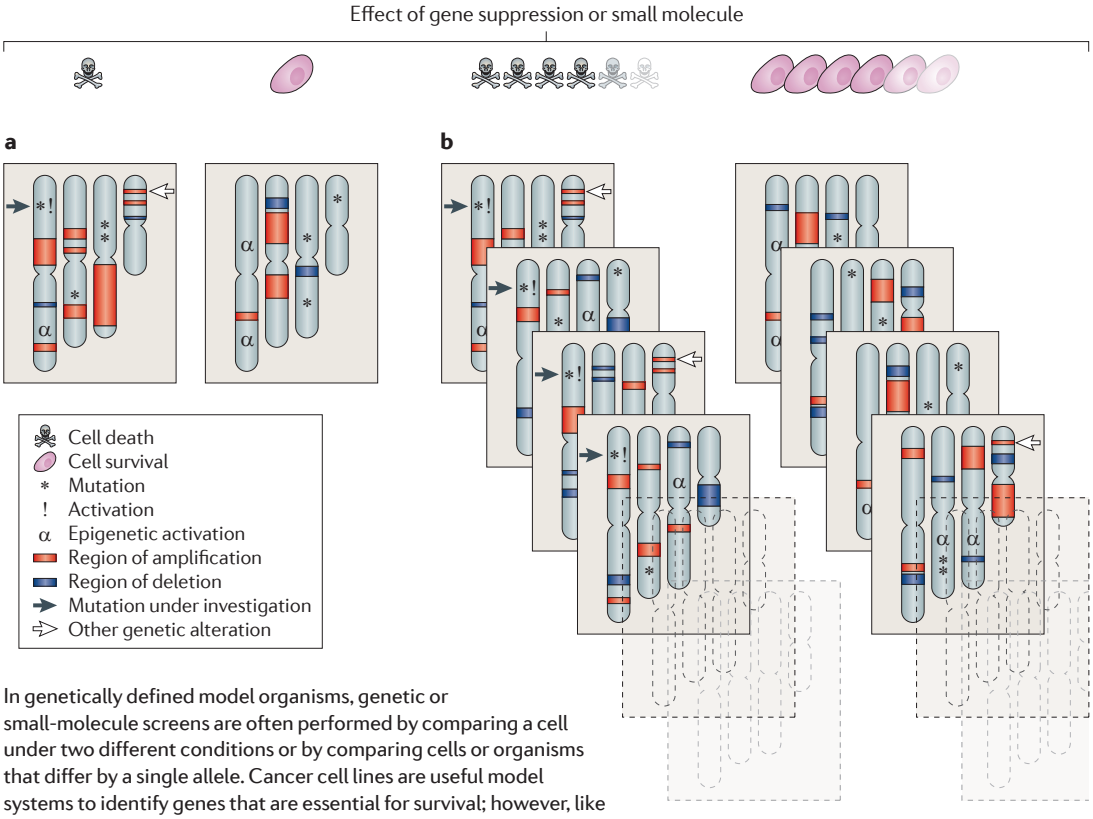
In each of the examples described above, genes were identified by their ability to confer a phenotype in which the surviving cells were easily scored. This experimental design is particularly well suited for the discovery of genes that mediate resistance to cytotoxic or targeted chemotherapy agents. Indeed, several investigators have used ORF and RNAi libraries to identify genes that are involved in the resistance to many agents. For example, using this approach, the PI3K pathway was identified as a resistance mechanism to a monoclonal antibody (trastuzumab) that targets the oncogene *ERBB2* (REF. 47). Expression of *RARA* or *PRAME* has been shown to mediate resistance to histone deacetylase (HDAC) inhibitors[71] and the kinases COT (also known as MAP3K8) and RAF1 have been shown to mediate resistance to BRAF inhibition[31]. Similarly,

Pooled screen
A functional-genomics screen in which genetic tools are mixed and administered to a cellular population under a selective pressure.

Aniokis
A form of cell death that is associated with loss of cell–matrix interactions.

RAS
A family of small GTPases that are frequently mutated in cancer. Single-nucleotide substitutions lead to constitutive activation of RAS signalling.

Box 2 | **Meta-analysis of functional screens**



In genetically defined model organisms, genetic or small-molecule screens are often performed by comparing a cell under two different conditions or by comparing cells or organisms that differ by a single allele. Cancer cell lines are useful model systems to identify genes that are essential for survival; however, like primary tumours, such cells harbour complex genotypes that pose a challenge to the interpretation of screens comparing pairs of cells.

The identification of genes or small molecules that are essential for cancer cells harbouring a specific genetic alteration (see the figure, part **a**, black arrow) could, theoretically, be accomplished by screening cancer or normal cells with and without the mutation. However, owing to the complexity of cancer genomes, mutations beyond the specific mutation investigated (white arrow) track with the observed phenotype. Thus, connecting genotypes to phenotypic observations that are derived from comparisons between small numbers of cancer cells is challenging.

Certain experimental designs will permit statistically powered meta-analyses of functional screens connecting genotypes with dependencies (see figure, part **b**). Because each screened cancer cell line harbours many alterations, a large panel of cell lines must be screened to identify dependencies that track with a genotype of interest. Furthermore, some cancer cells without a particular mutation may nonetheless have an activated protein product. Thus, performing such experiments in large panels of cancer cells provides sufficient power to overcome such misclassifications. This experimental design involves a larger investment to achieve the number of measurements necessary to make meaningful comparisons, and it requires the use of statistical methods to identify perturbagens that show selective lethality. Although analytical approaches that were developed for expression microarray analysis have been adapted to analyse these types of experiments[93], it is clear that new approaches are needed. In model organisms, screens are usually performed until genetic saturation occurs, at which point additional screening only identifies previously discovered candidates. In mammalian cells, achieving the same degree of genetic saturation will require further technical advances. The use of meta-analyses, however, will facilitate the identification of true positives.

loss-of-function screens have been used to identify genes whose suppression mediates resistance to chemotherapy drugs such as campthothecan[72], HDAC inhibitors[73], bortezimib[74] and lapatanib[75] and have been used to identify biomarkers that may predict responsiveness to experimental therapeutics[76].

In addition to these direct measurements, signatures that represent a particular cell state can be used to identify genes or small molecules that perturb gene function. For example, a limited number of genes represent a signature that distinguishes between leukaemic cells and differentiated mononuclear cells, and a bead-based system that can quantify the expression of these genes has been used to screen for small molecules that modulate this signature[77,78]. Also, an assay to measure the transcription of fetal haemoglobin identified HDAC1 and HDAC2 inhibitors as essential for the switch to adult haemoglobin[79]. As technologies that permit multiplexed monitoring of different types of cell signatures continue to mature, signature-based readouts, coupled to large databases of existing signatures and efficient pattern-matching algorithms[80], will permit the identification of novel relationships between individual genes, between genes and pathways, and between

Perturbagens
Small molecules, peptides, cDNAs or RNAi inducers that disrupt biological processes.

Box 3 | **Image-based readouts**

Although most genetic screens performed to date involve the measurement of proliferation or cell survival, the recent development of high-quality or high-throughput imaging devices is likely to facilitate the investigation of a wide range of cell phenotypes. Specifically, screens involving the use of imaging enable researchers to use antibodies or intracellular probes to interrogate changes in cell shape, protein localization or differentiation state. Moreover, as the data from these screens can be archived, the same images can be re-analysed to study multiple phenotypes simultaneously. Similarly, fluorescence-activated cell sorting (FACS) can be used to separate cells by the expression of specific molecules. Tools to analyse these types of screens are now available[62]. The limitations of this approach include the availability and throughput of current microscopes, as well as the cost of the reagents required.

genes and cancer-associated phenotypes. Signatures that represent particular cell states can include mRNA expression, protein interaction[81] and/or activity[82,83] and high-content imaging[62] (FIG. 1).

Applying functional approaches to cancer

Beyond proof-of-principle experiments that demonstrate the feasibility and performance of specific libraries and reagents, several laboratories have begun to apply large-scale approaches to investigate the consequences of manipulating gene expression in a diverse array of experimental systems. In this section, we review some of the functional approaches that have been used to identify cancer genes and putative targets and we describe the integration of these approaches with outputs from cancer genome profiling efforts.

Targeted screens for specific gene families. One advantage of targeting a limited group of genes is that doing so minimizes the size and complexity of screening experiments. For example, by targeting known and predicted kinases with RNAi, several groups have identified kinases that are involved in a diverse set of phenotypes, including mitosis[17], apoptosis[14], aniokis[61], Hedgehog signalling[84], nuclear factor-κB (NF-κB) signalling[85], ovarian cancer[86] and survival in the presence of the human papillomavirus protein E7 (REF. 87). Other groups have focused on other families of enzymes. By designing a library of shRNAs targeting de-ubiquitylating enzymes, cylindromatosis (*CYLD*) was shown to be a tumour suppressor gene that regulates NF-κB activity[88] and USP1 was discovered as a new component of the Fanconi anaemia pathway that regulates the monoubiquitylation of the Fanconi anaemia group D2 protein (FANCD2)[89]. As small-molecule inhibitors of some oncogenic kinases and other enzyme classes exhibit clinical efficacy in specific cancer types, these studies identify a number of potential therapeutic targets. However, further work will be necessary to assess whether pharmacologic inhibition of these targets will produce the same effects as genetic suppression, as well as to determine the context (or contexts) in which these enzymes act.

Similarly, several groups have created expression libraries of specific gene families for identifying genes that are involved in producing cancer phenotypes. For example, the non-canonical NF-κB pathway regulator inhibitor of κB kinase-ε (*IKBKE*) was identified as an oncogene amplified in up to 30% of primary breast cancers by screening for kinases that could transform human cells[90]. Furthermore, a library composed of active and inactive versions of most human kinases was used to identify the kinases *MAP3K10* and *DYRK2* as novel regulators of the Hedgehog signalling pathway[91]. Kinase expression libraries have also informed studies of resistance. For example, recent work has identified HSPB8 as a kinase that mediates resistance to tamoxifen[92] and *COT* overexpression as one mechanism of resistance to a small-molecule inhibitor of mutant BRAF[31]. The availability of more complete collections of ORFs will facilitate focused and increasingly unbiased screens to identify genes involved in cancer.

Screens for essential genes in specific cellular contexts. Although loss-of-function screens provide the means to identify essential genes in almost any assay, one particularly attractive application of this approach is in identifying genes that exhibit differential essentiality in particular contexts (BOX 4). Similar to synthetic phenotype screens that are routinely performed in non-mammalian model organisms, several investigators have used medium- and large-scale loss-of-function screens to identify genes whose expression is necessary for survival in specific genetic contexts. For example, several genes have been identified as essential in the context of cells dependent upon oncogenic *KRAS*, including the kinases *TBK1, STK33, SNAI2* and *PLK1*, as well as Wilms tumour 1 (*WT1*)[93–97]. Similarly, two kinases, *SGK2* and *PAK3*, were found to be essential in the setting of loss of p53 function[98], and the proto-oncogene *MET*, cyclin-dependent kinase 6 (*CDK6*) and *MEK1* were found to be essential for the survival of cells lacking the functional von Hippel-Lindau tumour suppressor (VHL)[99]. Suppression of *CDK5* was found to enhance poly(ADP-ribose) polymerase (PARP)-inhibitor-induced cell death[100], and interferon regulatory factor 4 (*IRF4*) was found to be essential in multiple myeloma[101]. These experiments support the notion that synthetic lethal partners of known oncogenes and tumour suppressors may be used to target tumours even when the partner oncogene or tumour suppressor gene is not amenable to pharmacological targeting. Indeed, by integrating results from systematic loss-of-function screens in various contexts with analyses of altered genes in cancer genomes, therapeutic targets can be identified and categorized by their potential contributions to cancer pathogenesis (FIG. 3).

Integrating multiple functional approaches. Although many laboratories have used gain- and loss-of-function approaches to identify genes involved in a wide array of cancer-associated phenotypes, the available tools that are used to manipulate gene function remain imperfect as both overexpression constructs and RNAi may induce off-target effects. Although the use of multiple cell lines and independent reagents can mitigate some of these false-positive results, another powerful approach to vet lists of genes that are derived from primary screens is by integrating several different data sets to identify

Synthetic lethal
A relationship between two genes in which the combined inactivation of the genes results in lethality, whereas the inactivation of either gene alone has no effect. It can also refer to a gene whose perturbation only results in lethality in the presence of a particular cellular feature (for example, mutation).

Box 4 | **Context-dependent non-oncogene essentiality**

Functional-genomics experiments have the capacity to identify genetic and pharmacologic perturbations that only suppress cellular viability in particular contexts. One such context that has been well-studied is oncogene addiction — the dependency of a tumour on a particular oncogene that is mutated in that tumour. However, other genotype or cell-state-dependency relationships are emerging. The identification of 'context-dependent' genes and small molecules is highly desirable because these could provide paths towards the development of therapeutics that can be rationally deployed in groups of patients whose tumours have that specific context. Here we describe specific illustrative examples of contexts in which the function of non-oncogenes is selectively required, as measured by cell survival.

Tumour lineage/cell-of-origin
Dependencies specific to most cancers in a particular lineage may correspond to relationships with genetic or epigenetic activation events of particular genes that are enriched within a tumour type, genetic requirements for essentiality in tumours sharing particular cells of origin, or both. Examples include the requirement for interferon regulatory factor 4 (*IRF4*) in multiple myeloma cells[101] and microphthalmia-associated transcription factor (MITF) in cutaneous melanoma[55].

Oncogene dependency
Some oncogenes, such as *KRAS*, are activated in tumours not only by mutations in the oncogene itself, but also by upstream genetic mechanisms (for example, mutations in particular receptor tyrosine kinases, or loss of neurofibromin 1 (NF1)). Recently, a number of secondary dependencies have been proposed (described in the main text) that are only relevant in the context of cancers exhibiting KRAS dependency, rather than *KRAS* mutation per se.

Oncogene or tumour suppressor gene mutation
To sustain particular oncogenic mutations, tumours may crucially depend on continued signalling from a secondary gene product that is rendered essential only in the context of the oncogenic mutation. This concept, which is termed synthetic lethality, has been clinically validated: tumours harbouring *BRCA1* or *BRCA2* mutations are critically dependent on poly(ADP-ribose) polymerase 1 (PARP1) function for base-excision repair[110,111] and are sensitive to the PARP inhibitor olaparib[112].

Chemically sensitized
Sublethal doses of chemotherapeutics or targeted therapies render cancer cells unusually dependent on particular genes. Such genes might represent useful combinatiorial targets. Examples include *ACRBP* and *TUBGCP2* dependency in paclitaxol-treated, non-small-cell lung cancer cells[15].

Cell state
Recent work suggests that cancer stem cells may harbour particular dependencies; treatment of such cells with salinomycin specifically led to cell death in comparison with differentiated cells[113]. Other cell states, such as the altered metabolic state of most cancer cells, may also provide targets.

genes involved in a specific phenotype. For example, we identified *IKBKE* as a breast cancer oncogene by using an overexpression screen to identify kinases that could replace AKT in the transformation of human cells and an RNAi screen to identify genes essential for the proliferation of breast cancer cells, in addition to using knowledge of genes located in regions of recurrent amplification in breast cancers[90]. Similarly, *CDK8* was identified at the intersection of hits from shRNA screens for genes involved in the regulation of β-catenin signalling and cell proliferation in colon cancer cells that were also recurrently amplified in colon cancer genomes[66].

Combining genetic and pharmacological approaches also provides a powerful approach for identifying pathways and genes involved in specific cancers. As described above, a gene-signature-based approach was used to identify small molecules that perturb specific gene signatures[78]. When used to identify genes that are involved in the differentiation of acute myeloid leukaemia (AML) cells, small-molecule epidermal growth-factor receptor (EGFR) inhibitors were found to modulate the signature. As EGFR was not expressed in the cell line used in this screen, an RNAi screen was used to search for kinases that also induced this gene expression change, leading to the identification of SYK[77]. A similar approach identified 3-phosphoinositide-dependent protein kinase 1 (PDK1) as a mediator of tamoxifen sensitivity[102]. Although these examples demonstrate the utility of integrating several approaches, it also clear that as such efforts expand to cover large numbers of cell lines and genes, new bioinformatic and statistical approaches will be needed to analyse and interpret these screens.

Integrating functional approaches with cancer genome characterization. Over the next several years, through the use of highly efficient, massively parallel sequencing technologies[1], international efforts will allow characterization the genomes and epigenomes of thousands of tumour samples and provide catalogues of genes that are mutated in specific cancers. One way to perform integrative focused screens is to target genes within a genomic region that shows recurrent copy-number gain in a specific malignancy. Although some regions of recurrent copy-number alteration harbour a single gene, most of these regions include several genes; therefore, coupling of structural- and functional-genomics studies is required. Several studies have used loss-of-function genetic reagents to identify essential genes that lie in amplified regions in human tumours. A subset of these genes is likely to be oncogenes to which cancer cells are 'addicted' (that is, they are reliant on them for survival). For example, the identification of *CDK8* as an essential gene in colon tumours harbouring amplifications in a region of chromosome 13 that contains 16 genes led to a deeper investigation of the essentiality of all genes in this region[66]. Suppression of each gene in this region by RNAi showed that *CDK8* was the only one required for the survival of colon cancer cells with chromosome 13 amplifications. Furthermore, loss-of-function approaches, combined with the analysis of somatic alterations in soft-tissue sarcomas, identified specific amplified and essential genes in particular subtypes of these tumours[103]. A complementary and more direct approach to identify oncogenes is to overexpress in model systems candidate genes that lie in amplified regions in human tumours. *IKBKE* was identified as a breast cancer oncogene by this approach[90] and a recent oncogenomic screen in hepatocellular carcinoma more broadly demonstrated the use of this approach by testing over 100 candidates *in vivo*, confirming fibroblast growth factor 19 (*FGF19*) as a novel liver cancer oncogene that is co-amplified with cyclin D1 (*CCND1*)[104].

Genes that lie in regions of hemizygous or homozygous deletion in human tumours can be rapidly interrogated by loss-of-function screens. In one example, shRNAs were used to target every gene in a region of chromosome 5q, hemizygous loss of which occurs in a subtype of myelodysplastic syndrome that is characterized by a defect

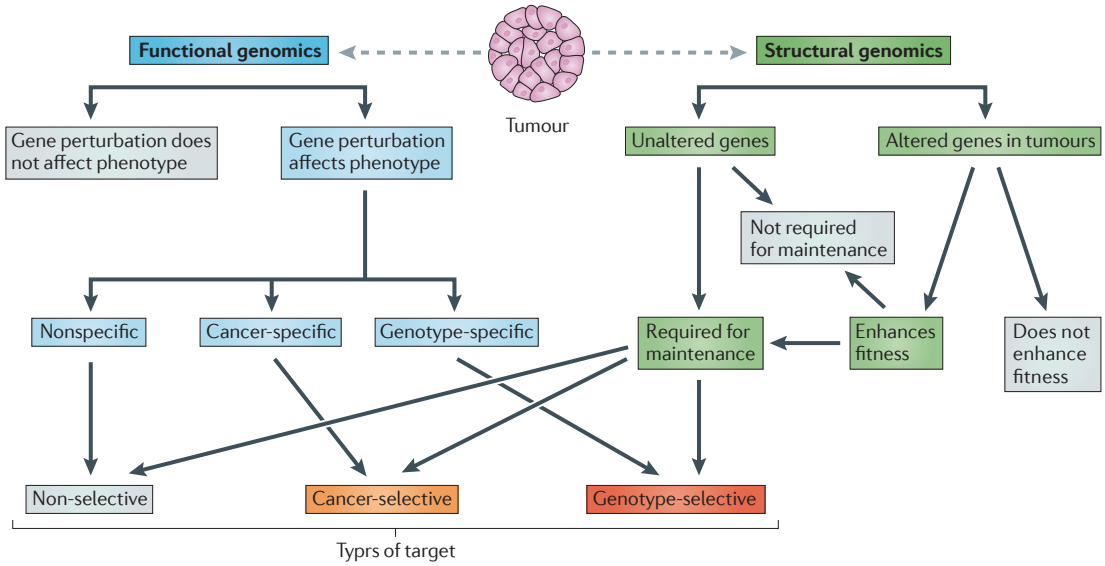


Figure 3 | **Integrating functional and structural cancer genomics.** An overarching goal of cancer genomics is to characterize alterations in tumour genomes (structural genomics) and associated phenotypic dependencies (functional genomics). Structural-genomics approaches distinguish tumour alterations predicted to have roles in increasing cancer fitness from other alterations that are present in tumours based primarily on the frequency that the mutation occurs. Distinguishing which mutations continue to be required for tumour maintenance, rather than tumour initiation only, requires integration with functional observations. In addition, unaltered genes may represent attractive targets as they may also be required for tumour maintenance. Functional-genomics approaches, such as loss-of-function screens, identify targets that are essential for all cells (cancer and normal), those that are selectively required for cancer cells and those that are selectively required for cancer cells harbouring particular genotypes.

in erythroid differentiation[105]. In normal haematopoietic progenitor cells, partial suppression of the ribosomal subunit protein RPS14 led to phenotypes found in this myelodysplastic syndrome subtype. In another example, RNAi reagents targeting deleted regions in human cancers were used to identify putative metastasis suppressor genes[106]. Furthermore, screens performed in an *ex vivo* format with shRNA reagents targeting genes in deleted regions in liver cancer and lymphoma have been highly successful in identifying a number of tumour suppressor genes in these disease types[56,57].

A complementary approach involves the analysis of syngenic regions of amplification or deletion in human cancer and tumours arising in murine experimental models. This integrated approach can help to assign priority to candidate genes for further functional investigation. For example, this strategy led to the identification of *YAP1* and *cIAP1* (also known as *BIRC2*) as hepatocellular oncogenes[107] and *NEDD9* as a gene involved in melanoma metastasis[108]. As genome characterization efforts of human and murine tumours proceed, similar approaches are likely to be useful in identifying the genes responsible for transformation and other cancer phenotypes in regions of recurrent copy-number change.

Although most integrated studies have, thus far, focused on regions of altered copy number in human tumours, the systematic functional characterization of mutated and translocated genes will require the generation and deployment of new ORF reagents to complement existing tools. Specifically, efforts to create mutant alleles will be required to test the function of such alleles in a battery of cancer phenotypes. Moreover, genes that are identified by unbiased functional screens are also candidates to interrogate for structural changes in tumours. For example, the finding that the myeloid primary response gene *MYD88* was differentially essential in the ABC subtype of diffuse large B-cell lymphoma in a large-scale RNAi screen prompted sequencing of *MYD88* in lymphoma samples and the identification of activating mutations in this gene[109].

Conclusions and future directions

The current pace of genome characterization will almost certainly provide comprehensive views of the genetic and epigenetic alterations that occur in cancer genomes. If combined with the tools and approaches described here (FIGS 1,2), it is increasingly likely that the major pathways involved in cancer development will be uncovered. In particular, by combining structural characterizations of cancer genomes with functional investigations, one should be able to identify somatically altered genes that contribute directly to tumour initiation and maintenance (FIG. 3), as well as context-specific dependencies (BOX 4). Indeed, early work using integrative approaches has not only identified new oncogenes but has also provided ways towards exploiting this knowledge therapeutically.

However, several areas require further development before comprehensive functional views of the cancer genome will be possible. First, despite progress in tool building, all of the available collections remain incomplete. Specifically, complete ORF or cDNA

Structural genomics
Genome-wide approaches for cataloguing structural changes (for example, mutations and copy-number changes) in the genome.

collections do not yet exist for mammalian genomes and alternative splice forms for individual genes remain poorly characterized. Similarly, further technical development is needed to permit loss-of-function approaches to achieve genetic saturation, and off-target effects of RNAi are not completely understood. In addition, because of the inherent heterogeneity in cancer and cancer models, these approaches will need to be applied at a much larger scale to obtain statistically meaningful conclusions.

Moreover, although proof-of-principle experiments have been reported using a diverse range of assays, most studies so far involve the investigation of cell survival as a primary readout. Further work is necessary to expand these studies to a much larger collection of cell lines and animal models and to perform similar studies in a much wider range of phenotypic assays, as well as using a diverse range of cellular signatures. In particular, assays that measure many aspects of cancer biology by using imaging or other readouts of cell signalling require further development. In parallel, assays that measure cell–cell interactions will be necessary to dissect such heterotypic interactions that are important for cancer development and to distinguish whether candidate genes function in a cell-autonomous or non-cell-autonomous manner. In addition, although the approaches used so far for functionally characterizing genes that are mutated in cancer genomes focus on tumour initiation, it is likely that other genes will have roles in tumour progression or metastasis. Other genes may drive or maintain tumorigenesis only in the context of particular cell types or together with specific cooperating sets of genetic alterations. Therefore, a major investment is warranted in the development and validation of progression and metastasis assays, as well as an expansion of the genetic backgrounds currently modelled by pre-malignant model systems.

The outputs of these types of experiments will provide the foundation for molecular, cellular and genetic approaches to investigate the role and function of specific genes and pathways in cancer biology. The information provided by systematic interrogation of cancer genomes will allow investigators to focus on deciphering the biology of crucial pathways, which is essential not only to understand the molecular basis of cancer but also to develop the means to target these pathways therapeutically. Although the approaches described in this article are focused on the application of these methods in studying cancer biology, they are not specific to this area and, as such, they provide proof-of-principle evidence supporting the feasibility of these approaches for the study of other biological phenomena and diseases.

1. Meyerson, M., Gabriel, S. & Getz, G. Advances in understanding cancer genomes through second-generation sequencing. *Nature Rev. Genet.* **11**, 685–696 (2010).
2. Maemondo, M. *et al.* Gefitinib or chemotherapy for non-small-cell lung cancer with mutated EGFR. *N. Engl. J. Med.* **362**, 2380–2388 (2010).
3. van de Vijver, M. J. *et al.* A gene-expression signature as a predictor of survival in breast cancer. *N. Engl. J. Med.* **347**, 1999–2009 (2002).
4. McDermott, U. *et al.* Genomic alterations of anaplastic lymphoma kinase may sensitize tumours to anaplastic lymphoma kinase inhibitors. *Cancer Res.* **68**, 3389–3395 (2008).
5. Soda, M. *et al.* Identification of the transforming *EML4-ALK* fusion gene in non-small-cell lung cancer. *Nature* **448**, 561–566 (2007).
6. Copeland, N. G. & Jenkins, N. A. Harnessing transposons for cancer gene discovery. *Nature Rev. Cancer* **10**, 696–706 (2010).
7. Starr, T. K. *et al.* A transposon-based genetic screen in mice identifies genes altered in colorectal cancer. *Science* **323**, 1747–1750 (2009).
8. Ding, S. *et al.* Efficient transposition of the *piggyBac* (*PB*) transposon in mammalian cells and mice. *Cell* **122**, 473–483 (2005).
9. Rad, R. *et al. PiggyBac* transposon mutagenesis: a tool for cancer gene discovery in mice. *Science* **330**, 1104–1107 (2010).
10. Sun, L. V. *et al.* PBmice: an integrated database system of *piggyBac* (*PB*) insertional mutations and their characterizations in mice. *Nucleic Acids Res.* **36**, D729–D734 (2008).
References 7–10 demonstrate the use of transposon-mediated insertional mutagenesis.
11. Kool, J. & Berns, A. High-throughput insertional mutagenesis screens in mice to identify oncogenic networks. *Nature Rev. Cancer* **9**, 389–399 (2009).
12. Uren, A. G. *et al.* Large-scale mutagenesis in p19(ARF)- and p53-deficient mice identifies cancer genes and their collaborative networks. *Cell* **133**, 727–741 (2008).
13. Carette, J. E. *et al.* Haploid genetic screens in human cells identify host factors used by pathogens. *Science* **326**, 1231–1235 (2009).
14. MacKeigan, J. P., Murphy, L. O. & Blenis, J. Sensitized RNAi screen of human kinases and phosphatases identifies new regulators of apoptosis and chemoresistance. *Nature Cell Biol.* **7**, 591–600 (2005).
15. Whitehurst, A. W. *et al.* Synthetic lethal screen identification of chemosensitizer loci in cancer cells. *Nature* **446**, 815–819 (2007).
16. Brummelkamp, T. R., Bernards, R. & Agami, R. A system for stable expression of short interfering RNAs in mammalian cells. *Science* **296**, 550–553 (2002).
17. Moffat, J. *et al.* A lentiviral RNAi library for human and mouse genes applied to an arrayed viral high-content screen. *Cell* **124**, 1283–1298 (2006).
18. Silva, J. M. *et al.* Second-generation shRNA libraries covering the mouse and human genomes. *Nature Genet.* **37**, 1281–1288 (2005).
19. Kittler, R. *et al.* Genome-wide resources of endoribonuclease-prepared short interfering RNAs for specific loss-of-function studies. *Nature Methods* **4**, 337–344 (2007).
References 16–19 describe genome-scale libraries of shRNAs.
20. Premsrirut, P. K. *et al.* A rapid and scalable system for studying gene function in mice using conditional RNA interference. *Cell* **145**, 145–158 (2011).
21. Fellmann, C. *et al.* Functional identification of optimized RNAi triggers using a massively parallel ensor assay. *Mol. Cell* **41**, 733–746 (2011).
22. Echeverri, C. J. *et al.* Minimizing the risk of reporting false positives in large-scale RNAi screens. *Nature Methods* **3**, 777–779 (2006).
23. Seed, B. & Aruffo, A. Molecular cloning of the CD2 antigen, the T-cell erythrocyte receptor, by a rapid immunoselection procedure. *Proc. Natl Acad. Sci. USA* **84**, 3365–3369 (1987).
24. Shih, C. & Weinberg, R. A. Isolation of a transforming sequence from a human bladder carcinoma cell line. *Cell* **29**, 161–169 (1982).
25. Lin, H. Y. *et al.* Expression cloning of an adenylate cyclase-coupled calcitonin receptor. *Science* **254**, 1022–1024 (1991).
26. Temple, G. *et al.* The completion of the Mammalian Gene Collection (MGC). *Genome Res.* **19**, 2324–2333 (2009).
27. Bechtel, S. *et al.* The full-ORF clone resource of the German cDNA Consortium. *BMC Genomics* **8**, 399 (2007).
28. Lamesch, P. *et al.* Human ORFeome 3.1: a resource of human open reading frames covering over 10,000 human genes. *Genomics* **89**, 307–315 (2007).
29. Ota, T. *et al.* Complete sequencing and characterization of 21,243 full-length human cDNAs. *Nature Genet.* **36**, 40–45 (2004).
30. Park, J. *et al.* Building a human kinase gene repository: bioinformatics, molecular cloning, and functional validation. *Proc. Natl Acad. Sci. USA* **102**, 8114–8119 (2005).
References 27–30 describe large-scale expression libraries.
31. Johannessen, C. M. *et al.* COT drives resistance to RAF inhibition through MAP kinase pathway reactivation. *Nature* **468**, 968–972 (2010).
32. Yang, X. *et al.* A public genome-scale lentiviral expression library of human ORFs. *Nature Methods* (in the press).
33. Griffiths-Jones, S. miRBase: the microRNA sequence database. *Methods Mol. Biol.* **342**, 129–138 (2006).
34. Calin, G. A. *et al.* Frequent deletions and downregulation of micro-RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. *Proc. Natl Acad. Sci. USA* **99**, 15524–15529 (2002).
35. Voorhoeve, P. M. *et al.* A genetic screen implicates miRNA-372 and miRNA-373 as oncogenes in testicular germ cell tumours. *Cell* **124**, 1169–1181 (2006).
36. McDermott, U. *et al.* Ligand-dependent platelet-derived growth factor receptor (PDGFR)-α activation sensitizes rare lung cancer and sarcoma cells to PDGFR kinase inhibitors. *Cancer Res.* **69**, 3937–3946 (2009).
37. McDermott, U. *et al.* Identification of genotype-correlated sensitivity to selective kinase inhibitors by using high-throughput tumour cell line profiling. *Proc. Natl Acad. Sci. USA* **104**, 19936–19941 (2007).
38. Schreiber, S. L. *et al.* Towards patient-based cancer therapeutics. *Nature Biotech.* **28**, 904–906 (2010).
39. Nielsen, T. E. & Schreiber, S. L. Towards the optimal screening collection: a synthesis strategy. *Angew. Chem. Int. Ed. Engl.* **47**, 48–56 (2008).
40. Stanton, B. Z. *et al.* A small molecule that binds Hedgehog and blocks its signalling in human cells. *Nature Chem. Biol.* **5**, 154–156 (2009).
41. Moellering, R. E. *et al.* Direct inhibition of the NOTCH transcription factor complex. *Nature* **462**, 182–188 (2009).
42. Zhu, S. *et al.* A small molecule primes embryonic stem cells for differentiation. *Cell Stem Cell* **4**, 416–426 (2009).
43. Wurdak, H. *et al.* An RNAi screen identifies TRRAP as a regulator of brain tumour-initiating cell differentiation. *Cell Stem Cell* **6**, 37–47 (2010).

44. Schmitz, M. H. *et al.* Live-cell imaging RNAi screen identifies PP2A-B55α and importin-β1 as key mitotic exit regulators in human cells. *Nature Cell Biol.* **12**, 886–893 (2010).
45. Wheeler, D. B., Carpenter, A. E. & Sabatini, D. M. Cell microarrays and RNA interference chip away at gene function. *Nature Genet.* **37**, S25–S30 (2005).
46. Vegas, A. J., Fuller, J. H. & Koehler, A. N. Small-molecule microarrays as tools in ligand discovery. *Chem. Soc. Rev.* **37**, 1385–1394 (2008).
47. Berns, K. *et al.* A functional genetic approach identifies the PI3K pathway as a major determinant of trastuzumab resistance in breast cancer. *Cancer Cell* **12**, 395–402 (2007).
48. Hurov, K. E., Cotta-Ramusino, C. & Elledge, S. J. A genetic screen identifies the Triple T complex required for DNA damage signalling and ATM and ATR stability. *Genes Dev.* **24**, 1939–1950 (2010).
49. Smogorzewska, A. *et al.* A genetic screen identifies FAN1, a Fanconi anaemia-associated nuclease necessary for DNA interstrand crosslink repair. *Mol. Cell* **39**, 36–47 (2010).
50. Paddison, P. J. *et al.* A resource for large-scale RNA-interference-based screens in mammals. *Nature* **428**, 427–431 (2004).
51. Brummelkamp, T. R. & Bernards, R. New tools for functional mammalian cancer genetics. *Nature Rev. Cancer* **3**, 781–789 (2003).
52. Ngo, V. N. *et al.* A loss-of-function RNA interference screen for molecular targets in cancer. *Nature* **441**, 106–110 (2006).
53. Buchholz, F., Kittler, R., Slabicki, M. & Theis, M. Enzymatically prepared RNAi libraries. *Nature Methods* **3**, 696–700 (2006).
54. Schlabach, M. R. *et al.* Cancer proliferation gene discovery through functional genomics. *Science* **319**, 620–624 (2008).
55. Luo, B. *et al.* Highly parallel identification of essential genes in cancer cells. *Proc. Natl Acad. Sci. USA* **105**, 20380–20385 (2008).
56. Zender, L. *et al.* An oncogenomics-based *in vivo* RNAi screen identifies tumour suppressors in liver cancer. *Cell* **135**, 852–864 (2008).
57. Bric, A. *et al.* Functional identification of tumour-suppressor genes through an *in vivo* RNA interference screen in a mouse lymphoma model. *Cancer Cell* **16**, 324–335 (2009).
58. Meacham, C. E., Ho, E. E., Dubrovsky, E., Gertler, F. B. & Haemann, M. T. *In vivo* RNAi screening identifies regulators of actin dynamics as key determinants of lymphoma progression. *Nature Genet.* **41**, 1133–1137 (2009).
59. Zuber, J. *et al.* Toolkit for evaluating genes required for proliferation and survival using tetracycline-regulated RNAi. *Nature Biotech.* **29**, 79–83 (2011).
60. Kittler, R. *et al.* Genome-scale RNAi profiling of cell division in human tissue culture cells. *Nature Cell Biol.* **9**, 1401–1412 (2007).
61. Cheng, H. *et al.* SIK1 couples LKB1 to p53-dependent anoikis and suppresses metastasis. *Sci. Signal.* **2**, ra35 (2009).
62. Carpenter, A. E. *et al.* CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol.* **7**, R100 (2006).
63. Berns, K. *et al.* A large-scale RNAi screen in human cells identifies new components of the p53 pathway. *Nature* **428**, 431–437 (2004).
64. Westbrook, T. F. *et al.* A genetic screen for candidate tumour suppressors identifies REST. *Cell* **121**, 837–848 (2005).
65. Kolfschoten, I. G. *et al.* A genetic screen identifies PITX1 as a suppressor of RAS activity and tumorigenicity. *Cell* **121**, 849–858 (2005).
66. Firestein, R. *et al.* CDK8 is a colorectal cancer oncogene that regulates β-catenin activity. *Nature* **455**, 547–551 (2008).
67. Peeper, D. S. *et al.* A functional screen identifies *hDRIL1* as an oncogene that rescues RAS-induced senescence. *Nature Cell Biol.* **4**, 148–153 (2002).
68. Gazin, C., Wajapeyee, N., Gobeil, S., Virbasius, C. M. & Green, M. R. An elaborate pathway required for Ras-mediated epigenetic silencing. *Nature* **449**, 1073–1077 (2007).
69. Collins, C. S. *et al.* A small interfering RNA screen for modulators of tumour cell motility identifies MAP4K4 as a promigratory kinase. *Proc. Natl Acad. Sci. USA* **103**, 3775–3780 (2006).
70. Smolen, G. A. *et al.* A genome-wide RNAi screen identifies multiple RSK-dependent regulators of cell migration. *Genes Dev.* **24**, 2654–2665 (2010).
71. Epping, M. T. *et al.* A functional genetic screen identifies retinoic acid signalling as a target of histone deacetylase inhibitors. *Proc. Natl Acad. Sci. USA* **104**, 17777–17782 (2007).
72. O'Connell, B. C. *et al.* A genome-wide camptothecin sensitivity screen identifies a mammalian MMS22L-NFKBIL2 complex required for genomic stability. *Mol. Cell* **40**, 645–657 (2010).
73. Fotheringham, S. *et al.* Genome-wide loss-of-function screen reveals an important role for the proteasome in HDAC inhibitor-induced apoptosis. *Cancer Cell* **15**, 57–66 (2009).
74. Chen, S. *et al.* Genome-wide siRNA screen for modulators of cell death induced by proteasome inhibitor bortezomib. *Cancer Res.* **70**, 4318–4326 (2010).
75. Eichhorn, P. J. *et al.* Phosphatidylinositol 3-kinase hyperactivation results in lapatinib resistance that is reversed by the mTOR/phosphatidylinositol 3-kinase inhibitor NVP-BEZ235. *Cancer Res.* **68**, 9221–9230 (2008).
76. Mullenders, J. *et al.* Candidate biomarkers of response to an experimental cancer drug identified through a large-scale RNA interference genetic screen. *Clin. Cancer Res.* **15**, 5811–5819 (2009).
77. Hahn, C. K. *et al.* Proteomic and genetic approaches identify Syk as an AML target. *Cancer Cell* **16**, 281–294 (2009).
78. Stegmaier, K. *et al.* Gene expression-based high-throughput screening (GE-HTS) and application to leukemia differentiation. *Nature Genet.* **36**, 257–263 (2004).
79. Bradner, J. E. *et al.* Chemical genetic strategy identifies histone deacetylase 1 (HDAC1) and HDAC2 as therapeutic targets in sickle cell disease. *Proc. Natl Acad. Sci. USA* **107**, 12617–12622 (2010).
80. Lamb, J. *et al.* The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* **313**, 1929–1935 (2006).
81. Du, J. *et al.* Bead-based profiling of tyrosine kinase phosphorylation identifies SRC as a potential target for glioblastoma therapy. *Nature Biotech.* **27**, 77–83 (2009).
82. Pujana, M. A. *et al.* Network modeling links breast cancer susceptibility and centrosome dysfunction. *Nature Genet.* **39**, 1338–1349 (2007).
83. Wilson, B., Liotta, L. A. & Petricoin, E. Monitoring proteins and protein networks using reverse phase protein arrays. *Disease Markers* **28**, 225–232 (2010).
84. Evangelista, M. *et al.* Kinome siRNA screen identifies regulators of ciliogenesis and hedgehog signal transduction. *Sci. Signal.* **1**, ra7 (2008).
85. Lam, L. T. *et al.* Compensatory IKKα activation of classical NF-κB signalling during IKKβ inhibition identified by an RNA interference sensitization screen. *Proc. Natl Acad. Sci. USA* **105**, 20798–20803 (2008).
86. Sheng, Q. *et al.* An activated ErbB3/NRG1 autocrine loop supports *in vivo* proliferation in ovarian cancer cells. *Cancer Cell* **17**, 298–310 (2010).
87. Baldwin, A. *et al.* Kinase requirements in human cells: II. Genetic interaction screens identify kinase requirements following HPV16 E7 expression in cancer cells. *Proc. Natl Acad. Sci. USA* **105**, 16478–16483 (2008).
88. Brummelkamp, T. R., Nijman, S. M., Dirac, A. M. & Bernards, R. Loss of the cylindromatosis tumour suppressor inhibits apoptosis by activating NF-κB. *Nature* **424**, 797–801 (2003).
This study provided the first demonstration that gene-family-focused screens can illuminate genes involved in a specific phenotype.
89. Nijman, S. M. *et al.* The deubiquitinating enzyme USP1 regulates the Fanconi anaemia pathway. *Mol. Cell* **17**, 331–339 (2005).
90. Boehm, J. S. *et al.* Integrative genomic approaches identify *IKBKE* as a breast cancer oncogene. *Cell* **129**, 1065–1079 (2007).
91. Varjosalo, M. *et al.* Application of active and kinase-deficient kinome collection for identification of kinases regulating hedgehog signalling. *Cell* **133**, 537–548 (2008).
92. Gonzalez-Malerva, L. *et al.* High-throughput ectopic expression screen for tamoxifen resistance identifies an atypical kinase that blocks autophagy. *Proc. Natl Acad. Sci. USA* **108**, 2058–2063 (2011).
93. Barbie, D. A. *et al.* Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* **462**, 108–112 (2009).
94. Luo, J. *et al.* A genome-wide RNAi screen identifies multiple synthetic lethal interactions with the Ras oncogene. *Cell* **137**, 835–848 (2009).
95. Scholl, C. *et al.* Synthetic lethal interaction between oncogenic KRAS dependency and suppression of STK33 in human cancer cells. *Cell* **137**, 821–834 (2009).
96. Vicent, S. *et al.* Wilms tumour 1 (WT1) regulates KRAS-driven oncogenesis and senescence in mouse and human models. *J. Clin. Invest.* **120**, 3940–3952 (2010).
97. Wang, Y. *et al.* Critical role for transcriptional repressor Snail2 in transformation by oncogenic RAS in colorectal carcinoma cells. *Oncogene* **29**, 4658–4670 (2010).
98. Baldwin, A. *et al.* Kinase requirements in human cells: V. Synthetic lethal interactions between p53 and the protein kinases SGK2 and PAK3. *Proc. Natl Acad. Sci. USA* **107**, 12463–12468 (2010).
99. Bommi-Reddy, A. *et al.* Kinase requirements in human cells: III. Altered kinase requirements in VHL–/– cancer cells detected in a pilot synthetic lethal screen. *Proc. Natl Acad. Sci. USA* **105**, 16484–16489 (2008).
100. Turner, N. C. *et al.* A synthetic lethal siRNA screen identifying genes mediating sensitivity to a PARP inhibitor. *EMBO J.* **27**, 1368–1377 (2008).
101. Shaffer, A. L. *et al.* IRF4 addiction in multiple myeloma. *Nature* **454**, 226–231 (2008).
102. Iorns, E., Lord, C. J. & Ashworth, A. Parallel RNAi and compound screens identify the PDK1 pathway as a target for tamoxifen sensitization. *Biochem. J.* **417**, 361–370 (2009).
103. Barretina, J. *et al.* Subtype-specific genomic alterations define new targets for soft-tissue sarcoma therapy. *Nature Genet.* **42**, 715–721 (2010).
104. Sawey, E. T. *et al.* Identification of a therapeutic strategy targeting amplified FGF19 in liver cancer by oncogenomic screening. *Cancer Cell* **19**, 347–358 (2011).
105. Ebert, B. L. *et al.* Identification of *RPS14* as a 5q– syndrome gene by RNA interference screen. *Nature* **451**, 335–339 (2008).
106. Silva, J. M. *et al.* Cyfip1 is a putative invasion suppressor in epithelial cancers. *Cell* **137**, 1047–1061 (2009).
107. Zender, L. *et al.* Identification and validation of oncogenes in liver cancer using an integrative oncogenomic approach. *Cell* **125**, 1253–1267 (2006).
108. Kim, M. *et al.* Comparative oncogenomics identifies *NEDD9* as a melanoma metastasis gene. *Cell* **125**, 1269–1281 (2006).
109. Ngo, V. N. *et al.* Oncogenically active MYD88 mutations in human lymphoma. *Nature* **470**, 115–119 (2011).
110. Bryant, H. E. *et al.* Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase. *Nature* **434**, 913–917 (2005).
111. Farmer, H. *et al.* Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature* **434**, 917–921 (2005).
112. Fong, P. C. *et al.* Inhibition of poly(ADP-ribose) polymerase in tumours from BRCA mutation carriers. *N.Engl. J. Med.* **361**, 123–134 (2009).
113. Gupta, P. B. *et al.* Identification of selective inhibitors of cancer stem cells by high-throughput screening. *Cell* **138**, 645–659 (2009).

Acknowledgements
We thank the members of the Hahn laboratory and the Broad Institute Cancer Program for discussions. W.C.H. is supported in part by grants from the US National Institutes of Health, the Starr Cancer Consortium, the Ivy Foundation, the H.L. Snyder Foundation and the Prostate Cancer Foundation. We apologize to those authors whose relevant work could not be cited owing to space considerations.

Competing interests statement
The authors declare competing financial interests: see Web version for details.

FURTHER INFORMATION
William C. Hahn's homepage: www.hahnlab.org
Cancer Program at the Broad Institute of MIT and Harvard: http://www.broadinstitute.org/scientific-community/science/programs/cancer

ALL LINKS ARE ACTIVE IN THE ONLINE PDF