Published in final edited form as:

Trends Biotechnol. 2011 November ; 29(11): 558–568. doi:10.1016/j.tibtech.2011.06.003.

Genome-wide Approaches for Cancer Gene Discovery

Paul M. Lizardi^{1,2,*}, Matteo Forloni¹, and Narendra Wajapeyee^{1,*}

¹Department of Pathology, Yale University School of Medicine, New Haven, CT 06520-8023, USA

²Yale Computational Biology and Bioinformatics

Abstract

One of the central aims of cancer research is to identify and characterize cancer-causing alterations in cancer genomes. In recent years, unprecedented advances in genome-wide sequencing, functional genomics technologies of RNA interference screens and methods to evaluate three-dimensional chromatin organization *in vivo* have resulted in important discoveries regarding human cancer. The cancer causing genes identified from these new genome-wide technologies have also provided opportunities for effective and personalized cancer therapy. In this review, we describe some of the most recent technologies for cancer gene discovery. We also provide specific examples where these technologies have proven remarkably successful in uncovering important cancer-causing alterations.

Introduction

While cancer research has generated an impressive body of knowledge about this highly diverse class of diseases, our understanding of neoplasia at the systems level remains in its infancy. Impressive advances in structural biology and protein interactomes are underway to establish the three-dimensional structures of key protein molecules with important roles in cancer, as well as defining the logic of protein interaction and modification networks. On the other hand, the field of genomics remains far from the lofty goal of defining cancer-specific chromatin structural logic and dynamics, and its relation to the abnormal transcriptional events that define disrupted regulatory states in cancer genomes. In this context, cancer genomics represents a daunting challenge, since cancer genomes are not static but instead exist in dynamic and flexible evolutionary trajectories, characterized by mutation and structural rearrangements, made even richer by lineage-specific epigenetic variegation and functional state mosaicism. Thus, a systems approach to cancer functional genomics demands dramatic improvements in the molecular analysis of tumor genomes and transcriptomes. Fortunately, the ongoing revolution in DNA sequencing and chromatin analysis technologies are up to the task, and are already providing the scientific community with an arsenal of powerful tools for meeting this challenge.

© 2011 Elsevier Ltd. All rights reserved

*Corresponding authors: Narendra Wajapeyee, Department of Pathology, Yale University School of Medicine, New Haven, CT 06520-8023, USA. Phone: 203-737-5070; FAX: 203-737-1064 Narendra.Wajapeyee@yale.edu and Paul M. Lizardi, Department of Pathology, Yale University School of Medicine, New Haven, CT 06520-8023, USA. Phone: 203-785-5017; FAX: 203-785-7303. Paul.Lizardi@yale.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Conflict of Interest Statement Authors declare no conflict of interest.

Here we outline and briefly review several examples of high throughput methodological approaches that have emerged over the last few years, which hold promise for rapid progress towards a systems view of genomics and transcriptomics in cancer cells. Most of these methodologies are already in use for cancer genome analyses and have led to the identification of multiple cancer-causing genes, genetic alterations and deregulated pathways. The availability of a reference human genome implied that DNA sequencing could become the main tool for the exploration of cancer genomes. High-throughput ChIP-seq and RNA-seq techniques are in principle capable of documenting most chromatin modification events and virtually all of the transcriptional events in a tumor cell. Many of the ongoing cancer genome projects incorporate the methods that we describe in this review (Box 1). Our aim is to provide the reader with a basic understanding of these contemporary, cutting-edge methodologies. In Table 1 we summarize the methods described in this review and provide some guidelines that will help researchers to identify a technology described in this review that may be most suitable for their research problem.

High-throughput sequencing of cancer genomes and their value for understanding cancer

During tumorigenesis normal cells undergo complex genetic and epigenetic changes to become cancerous [1] [1–3]. Earlier studies aimed at evaluating the role of cancer-causing genes used hypothesis-driven candidate gene-based approaches. These approaches relied on first sequencing the gene of interest, which led to the identification of mutations or mutation hotspots within a gene. This was followed by characterization of the identified mutation regarding its functional impact on the protein of interest, eventually implicating both the mutation and the functional alteration in tumorigenesis. Successes of these candidate-based approaches are well represented by the identification of mutations in the tumor suppressor genes p53 and PTEN [4–6]. Although targeted gene sequencing approaches are straightforward and inexpensive, they are biased and low throughput. Furthermore, due to the complex nature of human cancers, these candidate-based strategies fail to provide the complete landscape of all the genomic changes that takes place in cancer. Therefore, genome-wide sequencing methods have now become the preferred methods of choice for cancer genome analyses [7, 8]. In recent years several new studies have taken advantage of high-throughput deep-sequencing methodologies for human cancer profiling [8–15]. Massively parallel DNA sequencing methods provide opportunities to carry out genome-wide screening for point mutations, copy number variation and rearrangements on a single platform. For example, a study by Chiang et al. used high-throughput sequencing for high-resolution mapping of copy-number alterations. Identification of genome regions that show copy number variation is a powerful way to discover cancer-causing genes. This study analyzed three matched pairs of normal and cancer cell lines. In sum, authors showed that a collection of ~14 million aligned sequence reads from human cell lines has comparable power to detect events as the current generation of DNA microarrays [16]. However, massive parallel sequencing showed over a two-fold higher precision for localizing breakpoints, typically, to within a ~1 kb region [16].

Genome-wide deep sequencing approaches and many other approaches that we describe later in this review are extremely effective in their discovery power and are applicable to any human cancer. However, they require specialized support in order to perform the bioinformatics analyses of genomic data. Some sequencing platforms such as Illumina's genome analyzer has their own pipeline that provide semi-processed data, which then require further analyses on custom bioinformatics platforms. This eventually allows for the mapping of the sequence reads to genomic regions to gain insight into genetic changes associated with cancer cells. Therefore, before venturing into high-throughput sequencing-based approaches, it is of utmost importance to first establish bioinformatics methods for the

analysis of large amounts of sequencing data. A good way to familiarize oneself with high-throughput sequencing datasets is to download some of the previously published datasets from the National Center for Biotechnology Information (NCBI) maintained Gene Expression Omnibus (GEO) and to practice analysis of these datasets. This will allow the researcher to gain a preliminary understanding of the type of analysis that need to be performed for such research projects. Some of the genomic data analysis issues pertaining to cancer genomes are discussed in a recent review by Chin et al. [17]. We also refer readers to the table 2 that compares different sequencing platforms and genome-wide as well as targeted sequencing approaches.

Exome Sequencing to identify cancer-causing genes

The term exome refers to the complete set of regions of the genome corresponding to mature messenger RNA (mRNA), most of which codes for proteins. The human genome contains ~180,000 exons translating into ~30 megabases of DNA sequence, which, in total, represents ~1% of the total genomic DNA content. Since most of the genes that are implicated in human cancers are protein-coding genes, exome sequencing is expected to identify important cancer-associated changes. Because exome sequencing requires the sequencing of only ~1% of the genome, it is therefore less expensive compared to whole cancer genome sequencing. However, exome sequencing poses a technical challenge of first enriching the genomic DNA sequence corresponding to exons before sequencing. A brief account of the target enrichment methods that can be used in combination with deep sequencing technologies are presented in table 2 and Box 2 and have also been discussed by others [18–20].

Many groups have now adopted the approach of targeted exome sequencing for detecting cancer-associated mutations [13, 21, 22]. A good example of this approach for cancer gene discovery is presented by a recent study that performed exome sequencing on 8 ovarian clear cell carcinoma (OCCC) cancer samples [13]. This study identified two previously unreported ovarian cancer genes: tumor suppressor ARID1B, a known chromatin modifier and an oncogene PPP2R1A [13]. The results of this genome-wide exomic sequencing were further confirmed by analyzing 42 additional samples of ovarian clear cell carcinomas. In this larger sample set, ARID1A and PPP2R1A were found to be mutated in 57% and 7% of the ovarian cancer samples, respectively [13].

This study and other similar studies [13, 21] have shown that, similar to genome-wide sequencing, exome sequencing can also reveal yet-to-be-discovered features of cancer genome, and will further our understanding of the role of protein coding genes in tumorigenesis and tumor progression.

Genome-wide RNAi screens and cDNA library screens for functional genomic analyses in cancer

The discovery of RNA interference in *C. elegans* and its adaptation to mammalian cells has revolutionized the way the biological function of a gene can be explored [23–26]. In 2001, two groups developed retroviral vector-based RNAi libraries, which were capable of targeting several thousand human genes [27, 28]. Using shRNA libraries, these groups demonstrated that unbiased genome-wide RNAi screens can be performed in mammalian cells and that RNAi screens constitute an extremely powerful functional genomics method for cancer gene discovery and validation [27, 28]. Since then RNAi libraries with several new features have been developed, including siRNA-based and lentiviral vector-based shRNA libraries [29, 30]. Many viral vector-based libraries are also barcoded, which allows for the performance of a RNAi screen using pools of several thousand shRNAs. The results

can later be deconvoluted to identify shRNAs using either barcode microarrays or deep-sequencing technologies, such as the Illumina genome analyzer [30–32]. Several shRNA and siRNA libraries are now available for the scientific community to use (Table 3). Availability of these libraries has allowed for the performance of several genome-wide RNAi screens, which have led to the identification of many new cancer genes and pathways [33–37]. In Box 3 we provide more details regarding RNAi screens and discuss two classes of RNAi screens in this section [38–40]. The first category of RNAi screens are referred to as *in vivo* RNAi screens and can be performed using a mouse model of human tumorigenesis. Using such a RNAi screen, Zender et al. combined the cancer genomics data with RNAi and a mosaic mouse model of hepatocellular carcinoma and identified hepatocellular carcinoma tumor suppressor genes [38]. First, based on comparative genome hybridization (CGH) arrays of hepatocellular carcinomas, the authors short-listed genes that were deleted in human samples of hepatocellular carcinoma. Following the identification of these genes, pools of shRNAs, targeting the genes identified from the CGH arrays were generated. These shRNA pools were then used to evaluate the tumor suppressive function of the targeted genes in a mouse model of hepatocellular carcinoma. Using this approach, the authors identified 13 tumor suppressor genes, 12 of which were previously not implicated in cancer. This study is very important as it provides an *in vivo* platform for integrating cancer genome data with RNAi screens and mouse model of cancers. In the future, these types of screens will prove extremely useful for functionally characterizing the genetic information obtained through several ongoing large-scale cancer genome analysis initiatives.

The second category of RNAi screens is aimed at targeting cancer cells with activating RAS mutations [39, 40]. These RNAi screens are based on the principle of synthetic lethality and aim to develop new ways to target cancer by exploiting non-oncogenic addictions in cancer cells [41–43]. In the first synthetic lethality screen Luo et al. performed a genome-wide RNAi screen in isogenic DLD-1 cell lines that either had oncogenic KRAS or wild type KRAS. The cell lines were infected with 6 pools of ~13,000 shRNAs/pool in triplicate. The authors then analyzed the change in relative abundance of each shRNA by microarray hybridization to identify and compare the lethality signature of the KRAS mutant and the KRAS WT DLD-1 cell lines. This led to the identification of shRNAs that were specifically synthetic lethal to RAS mutant cells. Furthermore, the authors observed an enrichment of shRNAs targeting genes with mitotic functions leading to the identification of a pharmacologically tractable pathway involving Polo Like Kinase 1 (PLK1).

In another synthetic lethality screen for KRAS mutant cancers another group used the RNAi Consortium's (TRC) shRNA library [40]. The authors used 5024 shRNAs targeting 1011 human genes, including the majority of known and putative protein kinase genes and a selection of protein phosphatase genes and known cancer-related genes [40]. The screen was performed in eight human cancer cell lines representing five different tumor types as well as fibroblasts and immortalized human mammary epithelial cells (HMECs) [40]. Through this RNAi screen, the authors identified and functionally validated that STK33 is required for the survival and proliferation of KRAS mutant cancer cells.

The results of these studies showcase the power of unbiased genome-wide RNAi screens in identifying and validating cancer genes and identifying important genetic relationships among genes in cancer cells.

Complementary to RNAi screens, cDNA library-based screens (also known as ORFome screens) can also be performed to identify gain-of-function events in cancer. A brief description of ORFome screens and considerations for performing such screens are presented in Box 4. A good example of this screen is represented by a very interesting study carried out by Boehm et al [44]. The authors performed an ORFome screen to identify

kinases that can substitute for myristylated-AKT (myr-AKT) in transforming cells in cooperation with constitutively active MEK (MEK^{DD}). To do so, authors cloned 354 kinases and kinases related open reading frames (ORFs) as myristylated and Flag-tagged [44]. After cloning and confirming the activity of these kinases, the authors performed a screen with pools of 10 to 12 unique ORFs to identify activated kinases that could substitute myr-AKT and induce transformation in cooperation with activated MEK1. The pools that allowed for anchorage-independent growth were then deconvoluted by testing the ORFs individually. Using this approach the authors were able to identify four kinases that can cooperate with activated MEK1 to cause anchorage-independent growth. One of the kinase identified was IKBKE (IKK ϵ). Further analysis by the authors showed that IKK ϵ is amplified and consequentially overexpressed in breast cancer. Furthermore, it is required for breast cancer cell survival as knocking down of IKK ϵ leads to apoptosis induction. This study provides a framework for performing integrative genomics analyses to identify gain-of-function events [44]. Eventually, in combination with RNAi-based approaches, this kind of methodology will allow researchers to assess the functional consequences of cancer cell-associated changes identified through genome-wide analysis initiatives.

Non-canonical cancer genes emerging from alternative promoters and alternative splicing

Promoters are DNA elements that allow for the recruitment of RNA polymerase, leading to the establishment of a specific transcriptional state [45]. Human genes can have one or multiple promoters, many of which can act as alternative promoters [46, 47]. Alternative promoters not only provide the opportunity for differential gene expression, but can also influence the relative amounts of alternative transcripts produced, as well as their stability [46, 47]. Genome-wide promoter analyses studies have indicated that over 50% of human genes have at least one alternative promoter [48, 49]. Interestingly, disease-associated genes are more likely to be associated with more alternative promoters and tend to be differentially expressed [50, 51]. Furthermore, a recent study has indicated that cancer-related genes have on average two alternative promoters compared to 1.5 for other human genes [50], indicating that cancer cells may in part rely on the usage of alternative promoters for tumorigenesis and even for the maintenance of tumorigenic state.

A recent study of Hodgkin's lymphoma discovered that loss of the transcriptional repressor CBFA2T3 leads to reactivation of a LTR of mammalian apparent LTR retrotransposon (MaLR) family THE1B, which then functions as an alternative promoter. THE1B then leads to transcriptional upregulation of protooncogene Colony Stimulating Factor 1 Receptor (CSF1R) in Hodgkin's lymphoma cells, which is required for the survival of these cells [52]. This study underpins the importance of analyzing the cancer transcriptome in order to identify the emergence of cancer cell-specific alternative promoters and evaluate their role in cancer.

Several methodologies have been developed in the last few years to identify promoters at genome-wide scale and to quantify their specific usage [53–57]. Here we describe one of these methods and evaluate its ability to identify alternative promoters, with the possibility of using it for cancer-specific alternative promoter discovery [54]. The method used by Sun et al. relies upon RNA pol II ChIP followed by ChIP-sequencing [54]. The authors performed this analysis for five different mouse tissues that included brain, liver, lung, spleen and kidney. Through their analysis, the authors were able to identify 38,639 promoters, 12,270 of which were novel promoters, which included both protein-coding and non-coding genes. Furthermore, by identifying the RNA pol II-bound promoter(s) of each annotated gene in a given tissue, the authors observed that 37% of the protein-coding genes use alternative promoters. This study provides an opportunity to probe cancer genomes and

compare them to normal cell genomes in order to understand how cancer specific-promoters may drive tumorigenesis.

Similar to alternative promoters, misregulation of pre-mRNA splicing can lead to oncogenic changes in otherwise normal genes [58]. One of the best examples of alternative splicing converting a non-tumorigenic protein to an oncogenic protein was studied in reference to splicing of the pyruvate kinase gene [59, 60]. Surprisingly, cancer cells exclusively express the embryonic splice isoform M2 of pyruvate kinase [59, 61, 62]. Switching of the M1 isoform to the M2 isoform in cancer cells results in increased anaerobic respiration and reduced oxidative phosphorylation, thereby contributing to tumorigenesis [59]. This study highlights the importance of identifying tumor specific gene isoforms, understanding their role in tumorigenesis and exploiting them for developing targeted cancer therapies.

Several methods have been used to identify alternative splicing events at a genome-wide scale [63–65]. One approach that has yielded remarkable results is RNA-sequencing [63, 64]. A recent study used RNA-seq to obtain a snapshot of the human transcriptome [64]. This study was able to identify 25% more transcripts compared to microarray technology. The authors used the RNA-seq data set to elucidate alternative splicing and were able to identify 95% of the splicing events expected in their dataset. This study identified 4096 previously unknown splice junctions in 3106 genes that were unique to one cell type. Within a given cell type, using junction reads, alternative splicing was identified in 30% of the expected genes and exon skipping was found to be largely overrepresented. Most importantly, the authors were able to decipher complex patterns of alternative splicing. For instance, with the use of EIF4G1 coding for the eukaryotic translation initiation factor 4 gamma 1, the authors showed 12 alternative splicing junction in B cells, five of which had not been identified by previous studies. Therefore, RNA-seq will be the preferred method for cancer genome transcriptomics, especially for identifying previously undocumented alternative splicing patterns in cancer cells, with the possibility of targeting cancer-cell specific isoforms for cancer treatment.

Beyond the conventional cancer gene concept: Abnormal cancer-related chromatin looping

The cancer genome is very dynamic. During tumorigenesis and cancer progression the constantly evolving cancer genome drastically affects the transcription profile through both local chromatin changes and through major alterations in long-range genomic interactions [66, 67].

Several studies have indicated that long-range chromatin interactions might be important for regulating the expression of oncogenes and imprinting of cancer-related genes [66]. The evidence of long range interactions and its effect on the oncogene MYC expression came from a study that analyzed the role of an inherited 8q24 cancer risk variant rs6983267, which is significantly associated with increased cancer risk in many malignancies, including colon cancer. The 8q24 cancer risk variant rs6983267 is extremely intriguing because of its location in a gene desert region. The authors noted that the nearest gene is located ~335 kb telomeric from this region and identified this gene as the proto-oncogene MYC. This study demonstrated that the 8q24 variant region shows all the features that define an enhancer. Using chromosome conformation capture (3C) technology, the authors demonstrated that the long-range interaction of this 8q24 variant enhancer with the MYC locus contributes to MYC overexpression leading to enhanced risk for colorectal cancer [66][68].

These observations are very timely because in 2009 two new methods were developed that can be applied to identify long-range interactions on a genome-wide scale [69, 70]. Both

these methods are adaptation of 3C technology for high-throughput genome-wide long-range interaction analysis [68]. In the first study, the authors developed a method called Hi-C, which can probe the three-dimensional architecture of human genome through proximity-based ligation followed by massive parallel sequencing. Utilizing Hi-C, the authors were able to construct spatial proximity maps of the human genome at a resolution of 1 megabase [69]. The second study developed a method, which the authors described as chromatin interaction analysis by paired end sequencing (ChIA-PET) [70]. ChIA-PET incorporates chromatin immunoprecipitation (ChIP)-based enrichment, chromatin proximity ligation, paired end tags and high-throughput sequencing. Using ChIA-PET the authors developed a human chromatin interactome of oestrogen-receptor- α (ER- α) binding. This data was further corroborated with trimethylation of lysine 4 (H3K4me3) ChIP, RNA pol II ChIP and gene expression arrays [70]. In summary, all of these results indicated that long-range interactions play an important role in transcription regulation by ER- α [70].

Both the Hi-C and ChIA-PET methods provide researchers with the opportunity to understand and evaluate cancer-specific changes that takes place in the context of three-dimensional cancer genome and use this information to decipher cancer-specific transcriptional regulation and its role in tumorigenesis.

Integration of high-throughput chromatin structure and transcription start site mappings

All of the methodologies that we have described are capable of providing large data sets of genome-wide information regarding cancer genome structure, chromatin modification states, gene transcriptional outputs or function of the genes. Interestingly, more powerful insights can be generated via the integration of multiple data sets, using the genomic coordinates of each observation as a common frame of reference.

For example, an RNAi screen may identify a specific transcription factor that acts as a key oncogene and is overexpressed in the development of a certain type of cancer. ChIP data obtained with an antibody specific for the same transcription factor can be used to identify the occupancy of hundreds of different genomic loci by this transcription factor, while a separate set of ChIP experiments can identify multiple histone modifications associated with loci bound by the transcription factor. Additionally, a Cap Analysis of Gene Expression (CAGE) experiment can be performed to identify all transcription start sites (TSS) induced by the transcription factor in the same cancer samples. CAGE [71–73] is a high-throughput method for analysis of the nucleotide sequence present at the capped 5'-terminus of RNA molecules. Capped molecules are selected from total RNA using affinity capture, and the resulting material is used to make complementary DNA, followed by ligation of a specially designed DNA tag sequence. The tagged cDNA is then processed further to generate concatenated DNA molecules containing the first 20 bases present at the 5'-terminus of each of the original capped RNA molecules. Sequencing of concatenated DNA enables mapping of the first 20 bases of each RNA to the corresponding location in the human genome. The exact positions of thousands of TSS are thus defined in a single CAGE experiment. The power of this approach lies in its ability to identify changes in transcriptional profiles that involve genes as well as regulatory RNA loci, such as microRNAs, antisense RNAs, long non-coding RNA, and other RNA regulatory molecules that may be specific to a cancer cell. By integrating the chromatin-based data and the CAGE-TSS mapping data, informative patterns of transcriptional regulatory logic may begin to emerge that extend way beyond our current gene-centric view of molecular abnormalities in cancer biology [74].

Conclusions

Human cancer poses an increasing health care challenge due to the aging population, and has dramatic socio-economic implications [75]. However, the high prevalence and complexity of cancer has also accelerated the development of powerful research tools for understanding all complex human diseases [76, 77]. Our review highlights a new generation of technologies that are being rapidly developed and deployed as enabling tools for ambitious cancer functional genomics initiatives, and on a broader scale, to combat all human diseases.

As much as these new genomics methodologies are advancing our understanding of cancer functional genomics, they are also creating new challenges for intelligent integration of the large amount of information obtained. We are just getting started with the task of placing this information in the context of a systems-view of human cancer that will elucidate regulatory logic at a deeper level, facilitating the development of novel cancer treatment strategies. In the future the cancer field will benefit from conceptual integration utilizing structural genomics, proteomics, and metabolomics-based approaches. An integrative, systems biology view will be crucial for advances in cancer prevention and cancer treatment, the ultimate goal of all cancer research efforts.

Acknowledgments

We apologize to those whose work could not be cited due to the space limitation. NW and PML are members of Yale Cancer Center. NW is a Sidney Kimmel Scholar for Cancer Research and is supported by Yale Department of Pathology Start-up funds, Yale Liver Center Pilot Grant (NIDDK P30-34989) and AACR Career Development Award for Pediatric Cancer Research (10-20-03-WAJA). PML acknowledges support from grants 1 R21 CA116079-01 and 5 R01GM080242-03 from the National Institutes of Health. We thank Job Dekker for suggestions and Alex Liu for help with the preparation of the manuscript.

Glossary

Chromatin Immunoprecipitation (ChIP)	ChIP is a technique for the immunoprecipitation-based enrichment of a specific protein cross-linked to genomic DNA. ChIP has been adapted to perform genome-wide analyses of the occupancy of DNA binding proteins, leading to better understanding of proteins that regulate gene expression by binding to DNA.
Massive parallel Sequencing	Massive parallel sequencing is a powerful cell-free sequencing method that can read millions of bases of DNA in a few hours. This methodology is faster, more accurate and significantly less expensive compared to traditional methods like capillary electrophoresis.
ChIP-sequencing (ChIP-seq)	ChIP-sequencing is a high-throughput methodology that combines ChIP with massively parallel sequencing. The use of massively parallel sequencing to analyze DNA segments allows ChIP-seq to achieve greater and deeper coverage in a shorter time compared to the traditional ChIP technique.
Positive-selection RNAi screen	A gain-of-function or positive-selection RNAi screen refers to a screen in which cells containing a specific shRNA that endows them with a growth advantage survive and proliferate in a given assay. For example, a screen that seeks to identify genes that, upon knockdown, rescue the growth inhibitory

	effects of a tumor suppressor would be categorized as a positive selection screen.
Drop-out RNAi screen	In contrast to a positive-selection screens, dropout RNAi screens aim to identify loss of proliferation/survival as an end point. For example, an RNAi screen that aims to identify essential housekeeping genes within a given cell type would require a negative-selection approach.
Synthetic lethality RNAi screen	Synthetic lethality refers to a situation in which a genetic defect leads to cell lethality only in combination with a second genetic change, while neither of these genetic changes alone is lethal.
In vivo RNAi screens	Since most biological processes occur within the context of an organism and are the outcome of interactions between multiple cell types, significant effort has recently been devoted to applying the techniques used to perform <i>in vitro</i> RNAi screens to whole organisms and achieve results that might be relevant in an <i>in vivo</i> context. Setting up the RNAi screens <i>in vivo</i> is relatively less cumbersome in planarian worms and <i>C. elegans</i> than it is in a mice model.
RNA-sequencing (RNA-seq)	RNA-seq is a powerful methodology used to sequence cDNA generated from cellular mRNA in order to gain insight into the complete transcriptome of a given cell. RNA-seq provides much deeper coverage compared to microarray-based methodologies and is able to obtain up to a single base-resolution making RNA-seq an invaluable tool for determining the cancer-specific transcriptome.
Mosaic Mouse Models	Mosaic mouse models are murine models where genetically altered stem cells are retro-transplanted into recipient mice thereby producing clusters of mutated cells surrounded by normal counterparts. Developing a mosaic mouse model is much faster and less expensive compared to traditional knockout mice-generation methodologies. The overall benefit of using mosaic mouse models lies in their suitability for testing the cooperation between multiple genetic lesions.
Chromosome Conformation Capture (3C)	3C is a methodology used to analyze the spatial organization of chromosomes, providing researchers with an opportunity to evaluate the existence of long-range chromatin interactions. In past years, 3C has been adapted for genome-wide surveys and has led to the development of technologies like 4C, 5C, Hi-C and ChIP-PET.

References

1. Kinzler KW, Vogelstein B. Lessons from hereditary colorectal cancer. *Cell*. 1996; 87:159–170. [PubMed: 8861899]
2. Jones PA, Baylin SB. The epigenomics of cancer. *Cell*. 2007; 128:683–692. [PubMed: 17320506]
3. Ting AH, et al. The cancer epigenome--components and functional correlates. *Genes Dev*. 2006; 20:3215–3231. [PubMed: 17158741]

4. Takahashi T, et al. p53: a frequent target for genetic abnormalities in lung cancer. *Science*. 1989; 246:491–494. [PubMed: 2554494]
5. Hollstein M, et al. p53 mutations in human cancers. *Science*. 1991; 253:49–53. [PubMed: 1905840]
6. Li J, et al. PTEN, a putative protein tyrosine phosphatase gene mutated in human brain, breast, and prostate cancer. *Science*. 1997; 275:1943–1947. [PubMed: 9072974]
7. Metzker ML. Sequencing technologies - the next generation. *Nat Rev Genet*. 2010; 11:31–46. [PubMed: 19997069]
8. Ley TJ, et al. DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature*. 2008; 456:66–72. [PubMed: 18987736]
9. Mardis ER, et al. Recurring mutations found by sequencing an acute myeloid leukemia genome. *N Engl J Med*. 2009; 361:1058–1066. [PubMed: 19657110]
10. Pleasance ED, et al. A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature*. 2010; 463:191–196. [PubMed: 20016485]
11. Taylor BS, et al. Integrative genomic profiling of human prostate cancer. *Cancer Cell*. 2010; 18:11–22. [PubMed: 20579941]
12. Lee W, et al. The mutation spectrum revealed by paired genome sequences from a lung cancer patient. *Nature*. 2010; 465:473–477. [PubMed: 20505728]
13. Jones S, et al. Frequent mutations of chromatin remodeling gene ARID1A in ovarian clear cell carcinoma. *Science*. 2010; 330:228–231. [PubMed: 20826764]
14. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature*. 2008; 455:1061–1068. [PubMed: 18772890]
15. Verhaak RG, et al. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell*. 2010; 17:98–110. [PubMed: 20129251]
16. Chiang DY, et al. High-resolution mapping of copy-number alterations with massively parallel sequencing. *Nat Methods*. 2009; 6:99–103. [PubMed: 19043412]
17. Chin L, et al. Making sense of cancer genomic data. *Genes Dev*. 2011; 25:534–555. [PubMed: 21406553]
18. Mamanova L, et al. Target-enrichment strategies for next-generation sequencing. *Nat Methods*. 2010; 7:111–118. [PubMed: 20111037]
19. Turner EH, et al. Massively parallel exon capture and library-free resequencing across 16 genomes. *Nat Methods*. 2009; 6:315–316. [PubMed: 19349981]
20. Hodges E, et al. Genome-wide in situ exon capture for selective resequencing. *Nat Genet*. 2007; 39:1522–1527. [PubMed: 17982454]
21. Jones S, et al. Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. *Science*. 2008; 321:1801–1806. [PubMed: 18772397]
22. Harbour JW, et al. Frequent Mutation of BAP1 in Metastasizing Uveal Melanomas. *Science*. 2010
23. Fire A, et al. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature*. 1998; 391:806–811. [PubMed: 9486653]
24. Elbashir SM, et al. Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature*. 2001; 411:494–498. [PubMed: 11373684]
25. Hammond SM, et al. Post-transcriptional gene silencing by double-stranded RNA. *Nat Rev Genet*. 2001; 2:110–119. [PubMed: 11253050]
26. Chang K, et al. Lessons from Nature: microRNA-based shRNA libraries. *Nat Methods*. 2006; 3:707–714. [PubMed: 16929316]
27. Paddison PJ, et al. A resource for large-scale RNA-interference-based screens in mammals. *Nature*. 2004; 428:427–431. [PubMed: 15042091]
28. Berns K, et al. A large-scale RNAi screen in human cells identifies new components of the p53 pathway. *Nature*. 2004; 428:431–437. [PubMed: 15042092]
29. Silva JM, et al. Second-generation shRNA libraries covering the mouse and human genomes. *Nat Genet*. 2005; 37:1281–1288. [PubMed: 16200065]
30. Bassik MC, et al. Rapid creation and quantitative monitoring of high coverage shRNA libraries. *Nat Methods*. 2009; 6:443–445. [PubMed: 19448642]

31. Silva JM, et al. Profiling essential genes in human mammary cells by multiplex RNAi screening. *Science*. 2008; 319:617–620. [PubMed: 18239125]
32. Schlabach MR, et al. Cancer proliferation gene discovery through functional genomics. *Science*. 2008; 319:620–624. [PubMed: 18239126]
33. Gazin C, et al. An elaborate pathway required for Ras-mediated epigenetic silencing. *Nature*. 2007; 449:1073–1077. [PubMed: 17960246]
34. Gobeil S, et al. A genome-wide shRNA screen identifies GAS1 as a novel melanoma metastasis suppressor gene. *Genes Dev*. 2008; 22:2932–2940. [PubMed: 18981472]
35. Wajapeyee N, et al. Oncogenic BRAF induces senescence and apoptosis through pathways mediated by the secreted protein IGFBP7. *Cell*. 2008; 132:363–374. [PubMed: 18267069]
36. Palakurthy RK, et al. Epigenetic silencing of the RASSF1A tumor suppressor gene through HOXB3-mediated induction of DNMT3B expression. *Mol Cell*. 2009; 36:219–230. [PubMed: 19854132]
37. Sheng Z, et al. A genome-wide RNA interference screen reveals an essential CREB3L2-ATF5-MCL1 survival pathway in malignant glioma with therapeutic implications. *Nat Med*. 2010; 16:671–677. [PubMed: 20495567]
38. Zender L, et al. An oncogenomics-based in vivo RNAi screen identifies tumor suppressors in liver cancer. *Cell*. 2008; 135:852–864. [PubMed: 19012953]
39. Luo J, et al. A genome-wide RNAi screen identifies multiple synthetic lethal interactions with the Ras oncogene. *Cell*. 2009; 137:835–848. [PubMed: 19490893]
40. Scholl C, et al. Synthetic lethal interaction between oncogenic KRAS dependency and STK33 suppression in human cancer cells. *Cell*. 2009; 137:821–834. [PubMed: 19490892]
41. Hartwell LH, et al. Integrating genetic approaches into the discovery of anticancer drugs. *Science*. 1997; 278:1064–1068. [PubMed: 9353181]
42. Kaelin WG Jr. The concept of synthetic lethality in the context of anticancer therapy. *Nat Rev Cancer*. 2005; 5:689–698. [PubMed: 16110319]
43. Luo J, et al. Principles of cancer therapy: oncogene and non-oncogene addiction. *Cell*. 2009; 136:823–837. [PubMed: 19269363]
44. Boehm JS, et al. Integrative genomic approaches identify IKBKE as a breast cancer oncogene. *Cell*. 2007; 129:1065–1079. [PubMed: 17574021]
45. Maston GA, et al. Transcriptional regulatory elements in the human genome. *Annu Rev Genomics Hum Genet*. 2006; 7:29–59. [PubMed: 16719718]
46. Schibler U, Sierra F. Alternative promoters in developmental gene expression. *Annu Rev Genet*. 1987; 21:237–257. [PubMed: 2450521]
47. Ayoubi TA, Van De Ven WJ. Regulation of gene expression by alternative promoters. *FASEB J*. 1996; 10:453–460. [PubMed: 8647344]
48. Kimura K, et al. Diversification of transcriptional modulation: large-scale identification and characterization of putative alternative promoters of human genes. *Genome Res*. 2006; 16:55–65. [PubMed: 16344560]
49. Baek D, et al. Characterization and predictive discovery of evolutionarily conserved mammalian alternative promoters. *Genome Res*. 2007; 17:145–155. [PubMed: 17210929]
50. Davuluri RV, et al. The functional consequences of alternative promoter use in mammalian genomes. *Trends Genet*. 2008; 24:167–177. [PubMed: 18329129]
51. Liu S. Increasing alternative promoter repertoires is positively associated with differential expression and disease susceptibility. *PLoS One*. 2010; 5:e9482. [PubMed: 20208995]
52. Lamprecht B, et al. Derepression of an endogenous long terminal repeat activates the CSF1R proto-oncogene in human lymphoma. *Nat Med*. 2010; 16:571–579. 571p following 579. [PubMed: 20436485]
53. Singer GA, et al. Genome-wide analysis of alternative promoters of human genes using a custom promoter tiling array. *BMC Genomics*. 2008; 9:349. [PubMed: 18655706]
54. Sun H, et al. Genome-wide mapping of RNA Pol-II promoter usage in mouse tissues by ChIP-seq. *Nucleic Acids Res*. 2010

55. Barrera LO, et al. Genome-wide mapping and analysis of active promoters in mouse embryonic stem cells and adult organs. *Genome Res.* 2008; 18:46–59. [PubMed: 18042645]
56. Sandelin A, et al. Mammalian RNA polymerase II core promoters: insights from genome-wide studies. *Nat Rev Genet.* 2007; 8:424–436. [PubMed: 17486122]
57. Carninci P, et al. Genome-wide analysis of mammalian promoter architecture and evolution. *Nat Genet.* 2006; 38:626–635. [PubMed: 16645617]
58. David CJ, Manley JL. Alternative pre-mRNA splicing regulation in cancer: pathways and programs unhinged. *Genes Dev.* 2010; 24:2343–2364. [PubMed: 21041405]
59. Christofk HR, et al. The M2 splice isoform of pyruvate kinase is important for cancer metabolism and tumour growth. *Nature.* 2008; 452:230–233. [PubMed: 18337823]
60. Christofk HR, et al. Pyruvate kinase M2 is a phosphotyrosine-binding protein. *Nature.* 2008; 452:181–186. [PubMed: 18337815]
61. Mazurek S, et al. Pyruvate kinase type M2 and its role in tumor growth and spreading. *Semin Cancer Biol.* 2005; 15:300–308. [PubMed: 15908230]
62. Dombrackas JD, et al. Structural basis for tumor pyruvate kinase M2 allosteric regulation and catalysis. *Biochemistry.* 2005; 44:9417–9429. [PubMed: 15996096]
63. Richard H, et al. Prediction of alternative isoforms from exon expression levels in RNA-Seq experiments. *Nucleic Acids Res.* 2010; 38:e112. [PubMed: 20150413]
64. Sultan M, et al. A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. *Science.* 2008; 321:956–960. [PubMed: 18599741]
65. Gupta S, et al. Genome wide identification and classification of alternative splicing based on EST data. *Bioinformatics.* 2004; 20:2579–2585. [PubMed: 15117759]
66. Pomerantz MM, et al. The 8q24 cancer risk variant rs6983267 shows long-range interaction with MYC in colorectal cancer. *Nat Genet.* 2009; 41:882–884. [PubMed: 19561607]
67. Vu TH, et al. Loss of IGF2 imprinting is associated with abrogation of long-range intrachromosomal interactions in human cancer cells. *Hum Mol Genet.* 2010; 19:901–919. [PubMed: 20015958]
68. Dekker J, et al. Capturing chromosome conformation. *Science.* 2002; 295:1306–1311. [PubMed: 11847345]
69. Lieberman-Aiden E, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science.* 2009; 326:289–293. [PubMed: 19815776]
70. Fullwood MJ, et al. An oestrogen-receptor- α -bound human chromatin interactome. *Nature.* 2009; 462:58–64. [PubMed: 19890323]
71. Balwierz PJ, et al. Methods for analyzing deep sequencing expression data: constructing the human and mouse promoterome with deepCAGE data. *Genome Biol.* 2009; 10:R79. [PubMed: 19624849]
72. Kodzius R, et al. CAGE: cap analysis of gene expression. *Nat Methods.* 2006; 3:211–222. [PubMed: 16489339]
73. Shiraki T, et al. Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. *Proc Natl Acad Sci U S A.* 2003; 100:15776–15781. [PubMed: 14663149]
74. Mattick JS, et al. A global view of genomic information--moving beyond the gene and the master regulator. *Trends Genet.* 2010; 26:21–28. [PubMed: 19944475]
75. Greenberg ER, et al. Social and economic factors in the choice of lung cancer treatment. A population-based study in two rural states. *N Engl J Med.* 1988; 318:612–617. [PubMed: 2830514]
76. Xie Y, Minna JD. Predicting the future for people with lung cancer. *Nat Med.* 2008; 14:812–813. [PubMed: 18685594]
77. Meyerson M, et al. Advances in understanding cancer genomes through second-generation sequencing. *Nat Rev Genet.* 2010; 11:685–696. [PubMed: 20847746]

BOX 1: Some noteworthy initiative for cancer genome analyses and characterization

Cancer Genome Anatomy Project (CGAP) (<http://cgap.nci.nih.gov/cgap.html>)

National Cancer Institute's (NCI) CGA project aims to determine the gene expression profile of normal, precancerous and cancerous cells, with the goal of improved detection, diagnosis and treatment for all the patients. The resources generated by CGAP are available to a broad research community. Interconnected modules provide access to all CGAP data, bioinformatics analysis tools and biological resources, allowing the user to find “in silico” answer to biological questions.

Cancer Genome Characterization Initiative (CGCI) (<http://cgap.nci.nih.gov/cgci.html>)

CGCI aims to assess the utility of new genomics technologies in strategically characterizing a subset of genomic changes involved in different tumors. Research groups involved with the CGCI initiative make all of their data available through a publicly accessible database. CGCI incorporates genomic characterization methods including exome and transcriptome analysis using second-generation sequencing technologies to better understand the underlying genetic changes that leads to cancer.

The Cancer Genome Atlas (TCGA) (<http://cancergenome.nih.gov/>)

National Cancer Institute (NCI) and the National Human Genome Research Institute (NHGRI) launched TCGA project to create a comprehensive atlas of the genomic changes involved in more than 20 common types of cancer. These analyses will be carried out over next five years. This large-scale, high-throughput effort is being carried out by a network of more than 100 researchers at different organizations across the USA. The long-term aims of TCGA, similar to other cancer genome analyses initiatives are to improve the technologies for cancer diagnosis and develop new methods for treating cancers.

The International Cancer Genome Consortium (ICGC) (<http://www.icgc.org/>)

The main goal for ICGC is to perform comprehensive analysis on 50 different tumor types or subtypes in order to gain insight into the genomic, transcriptomic and epigenomic changes associated with these tumor types. 22 countries and 120 research groups are participating in the ICGC studies. The data generated through this initiative will be freely available to researchers.

Box 2: Methods for Target enrichments for Exome-sequencing**PCR-based methods for target enrichment**

PCR based techniques are by far the simplest and most straightforward ones for target enrichment. A multiplex PCR can be used for large scale target enrichment followed by sequencing. RainDance technologies (<http://www.raindancetechnologies.com>) has adapted the PCR-based method for target enrichment in preparation for massively parallel sequencing.

Molecular Inversion probes for target enrichment

Molecular inversion probes (MIPs) are single stranded DNA sequences and are complementary to the target regions. MIPs have linker that can be used for PCR amplification. For exon capture applications, MIPs can be used in combination with DNA polymerase-based gap filling, thereby covering the intermediate regions for complete coverage of the target sequences.

Microarray-based target enrichment methods

Microarray-based target enrichment methods depend on the nucleic acid hybridization principle. Solid-phase methods for target enrichment utilize DNA probes on a microarray platform. To perform solid-phase capture, first, the fragmented genome can be ligated to an adapter sequence, followed by hybridization and then elution. This leads to selective enrichment for the region of choice. Following this, the enriched genomic region can be amplified using PCR and used for deep sequencing of the enriched sequences.

In-solution target enrichment methods

In-solution target enrichment methods, similar to solid-phase microarray-based enrichment methods, depend on the nucleic acid hybridization principle. However, in-solution target enrichment provides an opportunity to increase the throughput, because it allows for the enrichment of a greater number of samples compared to microarray-based methods in the same amount of time. Also, this method of target enrichment is less resource-intensive and cost-effective solutions for deep sequencing methodologies.

Box 3: RNAi screens**siRNA versus shRNA-based screens**

Genome-wide RNAi provide an opportunity to perform unbiased loss-of-function screens. In general, RNAi screens utilize either small interfering RNAs (siRNAs) or short hairpin-based RNAs (shRNAs). siRNA-based screens are particularly suitable in cases where desired phenotypes or the outcome of an assay can be observed in a short time frame without requirement for several rounds of cell division. Example of these types of RNAi screen includes apoptosis regulator and transcription regulator identification. However, some experiments that require several rounds of cell division before a phenotype can be observed, stable knockdown using shRNAs may be required. Examples of such RNAi screens will include the identification of genes required for cellular transformation.

End Point Choices: Survival versus High-content screens

RNAi screens can be set up where readout is survival (example includes positive selection and drop-out screens). Depending on the research questions high-content RNAi screens might be more appropriate. High-content RNAi screens are usually set up in multi-well plates where cells are monitored for morphological, molecular or transcriptional changes using fluorescence or luminescence-based assays. Therefore, high-content screens should be performed using siRNAs targeting one gene per well of the multi-well plate in almost all occasions. It should be noted that at a genome-wide scale these screens can be quite expansive and will require a substantial automation for set up and end point monitoring.

Box 4 Gain-of-function screens using cDNA libraries

Gain of function screens using cDNA libraries (also called ORFome screens) are complementary to RNAi screens and allow researchers to assess gain-of-function phenotypes in cancers. However contrary to siRNA/shRNA libraries, there is no standard cDNA library that can be used to address all types of gain-of-function questions. For example, a standard HeLa mammalian cell expression cDNA library can not be used to assess gain of function phenotypes in lung adenocarcinoma with a specific gain of function mutation or with a transforming oncogene resulting from a fusion such as the EML4-ALK fusion gene.

Although, ORFome screens pose more challenges in terms of assessing a given tumor's gain of function changes, once set up they are as powerful as RNAi screen in their discovery potential. One of the most important considerations when setting up ORFome screen will be to first choose an appropriate tissue to generate cDNA library. Next, the cDNA libraries will need to be normalized to ensure that high-expressing and low-expressing transcripts are represented equally in the library thereby preventing the loss of low expressers due to underrepresentation. Finally, researchers trying to perform ORFome should realize that based on the size or abundance of the transcript, they may end up missing some of the gain-of-function events. However, shRNA screens also have the same drawback in that they might not be saturating due to lack of shRNAs against a gene or its inability to cause gene knockdown.

The figure describes different experimental approaches that can be used to gain genome-wide insight into the changes associated with cancer cells. Methodologies like RNAi helps in both identifying new cancer genes and functionally validating them as important regulator of tumorigenesis.

The figure describes different experimental approaches that can be used to gain genome-wide insight into the changes associated with cancer cells. Methodologies like RNAi helps in both identifying new cancer genes and functionally validating them as important regulator of tumorigenesis.

Table 1

Genome-wide methods for cancer gene discovery.

Method	Applications	Requirements	General considerations for use
Genome-wide deep sequencing	Suitable for comprehensive human cancer profiling applications, such as the identification of mutations, SNPs, deletions, amplifications, copy number aberrations (CNAs) and gene fusion events.	Requires large data storage space and specialized bioinformatics support.	For specific needs, less resource intensive alternative approaches can be explored. For example, SNP arrays can provide information on SNPs, CNAs, deletions and amplifications in the cancer genome. Compared to a genome-wide deep sequencing approach, SNP arrays are less expensive and require only limited bioinformatics support.
Exome sequencing	Suitable for targeted sequencing of coding regions of the genome. Similar to genome-wide deep sequencing, exome sequencing can identify mutations, SNPs, deletions, amplification, and CNAs.	Requires specific methods to enrich the target sequences (also see box 2 and table 2). Requires large data storage space and specialized bioinformatics support. Less resource intensive compared to the whole genome sequencing as only ~1% of the entire genome needs to be sequenced.	In cases where the researcher wants information only about the coding regions, this may be a method of choice. However if only a small number of genes are being analyzed (for example 50 genes) in multiple cancer samples, these can still be deep-sequenced after targeted PCR-based amplification and pooling of the PCR product. This approach may substantially reduce the time required for sample preparation. Also, this will be the most cost-effective approach both in terms of sample preparation, deep sequencing and data analyses.
Genome-wide RNAi screens	Suitable for functional analysis of any gene through a loss-of-function method. An RNAi methodology can be adapted to perform positive selection, negative selection and synthetic lethality screens. In recent years, mouse model-based <i>in vivo</i> RNAi screens have also been performed (see Box 3).	siRNA-based or shRNA-based high content screens may require robotics and automation. Limited robotics are required for pooling of shRNA screens. Drop out screens, synthetic lethality screens and <i>in vivo</i> screens may require deep-sequencing based approaches or barcode microarrays to deconvolute the results.	In several instances where preliminary results point to a specific pathway(s) of interest, small scale RNAi screens targeting a limited number of genes can be performed with limited resources to approach the biological problem.
ORFome screens	Suitable for evaluating gain-of-function events in cancer cells that have been identified by genome-wide or targeted sequencing approaches (see Box 4).	May require a specific cDNA library construction depending on the tumor type being analyzed and the biological question being posed.	ORFome screens are as powerful as siRNA/shRNA screens. However, they should be carefully designed to avoid loss of evaluation of rare oncogenic changes (also see Box 4)
ChIP-seq	Can be used for global DNA-protein interaction analyses. Specifically useful for unbiased transcription factor binding site identification, alternative promoter identification and for studying histone modifications.	Requires good quality antibodies for Chromatin Immunoprecipitation (ChIP). Requires large data storage space and strong bioinformatics support.	ChIP-on-ChIP, which requires limited bioinformatics support and resources, can be first explored as an alternative to ChIP-seq. With advances in array technologies, high-density arrays providing almost genome-wide coverage are now available and may suffice in many cases.
RNA-seq	Useful for the study of gene transcription regulation or RNA processing; Particularly suitable for identifying rare splicing events and alternative splicing and global changes in RNA editing.	Requires a specific pipeline for data analysis. Limited to transcribed sequences. May require special library preparation protocols in cases where both mRNAs and small RNAs need to be sequenced.	Although very useful for identifying splicing events and rare transcriptomics changes, RNA-seq should not be used as an alternative for gene expression arrays.
ChIA-PET	ChIP-based 3-dimensional conformation analysis. Useful for long-range interaction studies	Requires deep sequencing and a specialized pipeline for sequence analysis.	If there are indications of a particular genomic loci being involved in a long-range interaction, a more

Method	Applications	Requirements	General considerations for use
		Requires specific antibodies suitable for ChIP.	directed analysis such as 3C may suffice.
Hi-C	An advanced version of chromosome conformation capture (3C); Can detect global chromatin conformation in any given cell type. Unlike ChIA-PET, Hi-C is not antibody based and therefore completely unbiased.	Requires a specific bioinformatics analysis pipeline.	If there are indications of a particular genomic loci being involved in long-range interaction, a more directed analysis such as 3C may suffice.
CAGE	Can be used to map transcription start sites (TSS) induced by a transcription factor in the same cancer samples.	Require strong bioinformatics support	Should not be used as a replacement for targeted strategies such as Rapid Amplification of cDNA Ends (5'-RACE) in cases where only few genes are being analyzed.

Table 2
Comparison of whole-genome sequencing and targeted sequencing approaches.

Whole-genome sequencing			General features		Targeted sequencing		
None			Sequence requirements		Prior knowledge of the target regions		
Entire genome			Coverage		Selected genomic regions of interest		
~ \$5 K– \$10 K			Sequencing cost		For exome sequencing cost is around 15% of that of sequencing the whole genome		
Decreased sample preparation time and complete full-genome coverage/Higher cost of sequencing			Advantages/disadvantages		Reduced sequencing costs and sequencing time/ Increased sample preparation cost and time		
454 Roche	Illumina	SOLID	Platform		In situ exon capture	MIP-based exon capture	PCR-based
Pyrosequencing	Solid-phase amplification	Sequencing-by-ligation chemistry	System		Array-based target capture	In-solution capture by Molecular Inversion Probes	PCR-based target amplification
100 Mb	1–1.5 Mb	1–4.5 Mb	Throughput ^a		On average of 2–3 Mb; up to 34 Mb each array with the HD2 NimbleGen array ^b	> 55,000 loci in a single assay by multiplex MIP	Hundreds to thousands of genomic loci in a single tube with the RainStorm platform ^b
~ \$5 K– \$6 K	~ \$5 K– \$6 K	~ \$5 K– \$10 K	Cost ^a		Medium	<10 samples, high; >100 samples, low	High
De novo genome sequencing	RIP-seq, ChIP-seq, transcriptome sequencing, expression profiling	Resequencing, expression profiling, structural rearrangement analysis with paired-end, large insert libraries.	Applications		Useful when there are a large number of genomic targets but a low number of samples	Useful when there are a large number of genomic targets and many samples	Useful for small/medium-scale studies
8 h	3 or more days	3 or more days	Run time ^a	Multiplex level ^c	10 ⁵ –10 ⁶	10 ⁴ –10 ⁵	10 ² –10 ³
200–300 nt	30–40 nt	30–40 nt	Read length	Sensitivity	~ 98.6%	>98%	~ 95%
~ 0.2% error rate	< 1.5% error rate	~ 0.2% error rate	Accuracy	Advantages/Disadvantages	High level of multiplexing/Limited resolution and specificity and high DNA input requirement (10–15ug)	Direct sequencing without the need for shotgun library construction, high specificity and low DNA input requirement (200ng)/Low capture uniformity	Potentially compatible with any next generation sequencing platform/ Individual oligonucleotide synthesis and large numbers of amplification reactions

^atime required for single run

^bRainStorm platform: <http://www.raindancetechnologies.com>; HD2 NimbleGen array (Roche): <http://www.nimblegen.com/products/>

^cNumber of probes used for target enrichment in each assay

Table 3
List of human and mouse shRNA/siRNA libraries available for genome-wide RNAi screens.

Library	Supplier/Institute	Genes Targeted	Total shRNA or siRNA	shRNA or siRNA/gene	Targeted Organisms	Vector	Type of Library	Features
pSM2 Retroviral Library	Thermo Scientific	28,000 mouse genes/28,500 human genes	~61,000/81,500	~3	Human and mouse	pSM2	Retroviral	miRNA-based design; Puromycin selection marker
GIPZ Lentiviral Library	Thermo Scientific	"Entire mouse genome"/"Entire human genome"	62,000/62,000	~2	Human and mouse	pGIPZ	Lentiviral	RNA pol II promoter; Turbo GFP; Puromycin selection marker; Can infect non-dividing cells
TRIPZ Inducible shRNA Library	Thermo Scientific	~16,000 human annotated genes/ ~15,950 mouse annotated genes	~159,000	~4-5	Human and mouse	pLKO.1	Lentiviral	Human U6 promoter; Inducible; Puromycin; Can infect non-dividing cells
MISSION shRNA Library	Sigma-Aldrich	~16,000 human annotated genes/ ~15,950 mouse annotated genes	~159,000	~4-5	Human and mouse	pLKO.1	Lentiviral	Human U6 promoter; Puromycin; Can infect non-dividing cells
NKI Library	NKI	~8,000 human genes/15,000 mouse genes	24,000/30,000	~2	Human and mouse	pRSC	Retroviral	RNA Pol III promoter; Puromycin selection
GeneNet shRNA Library	System Biosciences	39,000 mouse genes/47,400 human genes	150,000/200,000	4	Human and mouse	HIV and FIV-based	Lentiviral	Fluorescent proteins such as GFP etc.; Puromycin selection
siGenome SMARTpool siRNA Library	Thermo Scientific (Dharmacon)	18,236 human genes		~4	Human		siRNA-based	Four mRNA regions targeted at once to reduce false negatives;

Library	Supplier/Institute	Genes Targeted	Total shRNA or siRNA	shRNA or siRNA/gene	Targeted Organisms	Vector	Type of Library	Features
Silencer siRNA Libraries	Ambion	12,585 human genes/11,134 mouse genes	37,755/33,402	~3	Human and mouse		siRNA-based	Guaranteed 75% silencing Can be used in low concentration (>30nM); Off-target, polymorphic, and antiviral inducing regions have been eliminated