Functional Genomic Screening Approaches in Mechanistic Toxicology and Potential Future Applications of CRISPR-Cas9

Hua Shen*, Cliona M. McHale*, Martyn T Smith, and Luoping Zhang**
Superfund Research Program, Division of Environmental Health Sciences, School of Public Health, University of California, Berkeley, California 94720, USA

Abstract

Characterizing variability in the extent and nature of responses to environmental exposures is a critical aspect of human health risk assessment. Chemical toxicants act by many different mechanisms, however, and the genes involved in adverse outcome pathways (AOPs) and AOP networks are not yet characterized. Functional genomic approaches can reveal both toxicity pathways and susceptibility genes, through knockdown or knockout of all non-essential genes in a cell of interest, and identification of genes associated with a toxicity phenotype following toxicant exposure. Screening approaches in yeast and human near-haploid leukemic KBM7 cells, have identified roles for genes and pathways involved in response to many toxicants but are limited by partial homology among yeast and human genes and limited relevance to normal diploid cells. RNA interference (RNAi) suppresses mRNA expression level but is limited by off-target effects (OTEs) and incomplete knockdown. The recently developed gene editing approach called clustered regularly interspaced short palindromic repeats-associated nuclease (CRISPR)-Cas9, can precisely knock-out most regions of the genome at the DNA level with fewer OTEs than RNAi, in multiple human cell types, thus overcoming the limitations of the other approaches. It has been used to identify genes involved in the response to chemical and microbial toxicants in several human cell types and could readily be extended to the systematic screening of large numbers of environmental chemicals. CRISPR-Cas9 can also repress and activate gene expression, including that of non-coding RNA, with near-saturation, thus offering the potential to more fully characterize AOPs and AOP networks. Finally, CRISPR-Cas9 can generate complex animal models in which to conduct preclinical toxicity testing at the level of individual genotypes or haplotypes. Therefore, CRISPR-Cas9 is a powerful and flexible functional genomic screening approach that can be harnessed to provide unprecedented mechanistic insight in the field of modern toxicology.

© 2015 Elsevier B.V. All rights reserved.

*Address correspondence to Dr. Luoping Zhang, School of Public Health, University of California–Berkeley, 388 Li Ka-Shing Center, Berkeley, CA 94720 USA. Phone: (510) 643-5189. Fax: (510) 642-0427. luoping@berkeley.edu.
**These authors contributed to the work equally.

8. Conflict of interest statement
The authors declare that there are no conflicts of interest.

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.
1. Introduction

1.1 Importance of understanding human susceptibility to chemical exposures

Humans vary in their susceptibility to toxicants. Current risk assessment approaches rely on uncertainty factors to account for inter-individual variation, which can greatly under- or over-estimate the risk for an individual or population. As genetic variation likely accounts for a significant proportion of these individual differences, an increased understanding of the genetic variability of toxicant response will enable more accurate chemical exposure risk assessment.

Currently, the main approaches to identifying gene-environment interactions in toxicant mediated disease are genome-wide association studies (GWAS) and candidate gene association studies, both of which have limitations. These methods examine associations between all variants in the human genome, or variants in sets of genes or pathways previously implicated in toxicant response, with a phenotypic outcome related to toxicant exposure. GWAS need to be conducted in large exposed and control populations in order to find significant associations among the large number of variants tested, which makes these studies expensive and often not feasible. The candidate gene association study approach, requires smaller study populations than GWAS, but is limited by incomplete understanding of modes of action, and thus relevant genes and pathways, for many toxicants. Thus, using existing techniques to identify all the chemical susceptibility genes will take many years; in the meantime, alternative approaches are needed to identify and prioritize genes for candidate gene association testing.

1.2 Identifying all toxicity pathways by functional genomics

In the contest of risk assessment, understanding mechanisms of action of toxicants in a broader systems biology context is important to understanding new information and adopting next generation risk assessment applications [1–7]. In the field of toxicology, mechanistic information is often described by simplified models called modes of action (MOAs) or adverse outcome pathways (AOPs), and by more complex models called AOP networks. The latter, called networks to indicate the interconnectedness of disease-causing AOPs, reflect the various biological pathways and mechanisms through which chemicals cause toxicity and disease. Elucidating these mechanisms and their inter-relatedness can be challenging. A genome-wide systems biology approach is needed to identify non-obvious pathways to toxicity and to understand how normal network function is altered following exposure to chemicals or stressors. Functional genomics offers just such an approach. Functional genomics uses omic data to describe gene (and protein) functions and interactions on a genome-wide scale using high-throughput methods.
1.3 The role of functional genomics in toxicity testing: evolution of the technologies

In the context of toxicology, functional genomic screening involves systematic knockdown or knockout of non-essential genes or proteins on a pathway or genome-wide basis, in a cell of interest, and measurement of a toxicity phenotype, such as lethality, viability or fitness, following exposure to various doses of a toxic compound. As it directly measures the phenotype, it informs the association of a specific gene and its corresponding loss-of-function in the cellular response to a compound [8, 9]. Thus, functional genomic screening has the potential to identify genes that confer resistance or susceptibility to toxicants and to reveal toxicity pathways, through pathway analysis. Through the application of now affordable high-throughput screening technologies, functional genomic screening provides a means to simultaneously and efficiently screen thousands of compounds in cell-based systems [10]. Thus, the approach will provide mechanistic insight across many genes, pathways and chemicals, supporting the definition of more specific in vitro toxicological endpoints and the development of targeted cell-based assays [11, 12], that ultimately will have better predictive power for adverse health effects in humans than do traditional animal toxicological studies.

Functional genomic screening has been conducted in budding and fission yeast, fruit flies, worms, and human cell lines using various techniques. In this review, we particularly discuss genomic screens using in vitro models such as yeast and haploid eukaryotes and tools such as RNA interference (RNAi) and the most recently developed clustered regularly interspaced short palindrome repeats-associated nuclease (CRISPR)-Cas9 gene editing system. This review aims to describe the main in vitro functional genomic screening approaches that have been developed and to discuss their advantages and limitations (summarized in Table 1) in the context of toxicity testing.

As discussed in detail in the following sections, each approach has its own “pros” and “cons”. For example, high-throughput yeast screening has been well established due to the availability of mutant clones for all non-essential yeast genes, but incomplete homology with mammalian genes and functions limits its relevance to human toxicity. To overcome this weakness, approaches using mammalian haploid cells such as the KBM7 human bone marrow cell line and mouse embryonic stem cells (ESCs), have been developed. However, the leukemic nature of KBM7 limits its relevance to toxicity in normal human cells. RNA interference (RNAi) and the most recently developed clustered regularly interspaced short palindrome repeats-associated nuclease (CRISPR)-Cas9 genome editing system overcome many of the limitations associated with the other previously developed approaches. They are both applicable to any transfectable cell type and organism, and CRISPR, in particular, is relatively easy to design and perform, is highly specific, efficient and well suited for high-throughput and multiplexed gene editing, and can produce precisely targeted knock-out of most regions of the genome at the DNA level [13]. Broad application of this novel CRISPR-Cas9 system has been recently proposed in many research areas, including gene-therapy, drug development, and understanding phenotype associated with genetic variations [14].

The goal of this review is to highlight the potential application of CRISPR-Cas9 application in toxicity screening of environmental toxicants and discovery of mechanism and susceptibility.
2. Functional genomics in yeast

2.1 Yeast screening process

For several reasons, the eukaryotic budding (*Saccharomyces cerevisiae*) and fission (*Schizosaccharomyces pombe*) yeasts are ideal models in which to conduct functional genomic screening studies [15]. First, yeast is a well-established and widely used eukaryotic model for molecular and cellular biology studies [16]. Second, yeast is a unicellular non-pathogenic organism with rapid and stable growth in both diploid and haploid status. Third, the genetic makeup of yeast is fully characterized and genetic manipulation is straightforward and easily implemented. In contrast to other higher eukaryotes, most of the yeast genes do not contain introns, which simplifies the process of computer-based gene identification when conducting genome-wide analyses [17]. Functional information is available for nearly every gene in yeast [15, 18]. Fourth, yeast genes are highly conserved in human cells and other higher eukaryotes. Nearly half of the human genes implicated in heritable diseases have yeast homologues [19]. Although yeast does not possess as many physiological mechanisms of response to cytotoxic compounds as do other organisms, many of the basic mechanisms in response to chemical and environmental stresses are apparently conserved between yeast and other higher eukaryotic organisms [20, 21].

Phenotype-based analysis in mutant strains is a powerful way to determine the role of genes in response to a toxicant on a gene-by-gene or genome-wide basis. Many approaches to generating mutant strains have been developed including genetic footprinting [22] random mutagenesis [23] and polymerase chain reaction (PCR) based gene disruption [24]. Genetic footprinting and random mutagenesis are both untargeted and the process of matching phenotypes to genes is slow. A novel PCR-based gene disruption strategy was developed to generate a deletion (null mutation) in each of the open reading frames (ORFs) in the yeast genome [24]. Each yeast gene can be deleted in a directed manner by replacing an individual ORF with a selectable marker (e.g. antibiotic resistance) linked to a 20-base molecular tag (or barcode) that functions as a unique strain identifier [25]. The use of molecular tags allows the mutant strains to be pooled and assayed for the phenotypes in parallel in a process called parallel deletion analysis (PDA) [26, 27].

2.2 Parallel deletion analysis (PDA)

The screening process is outlined in Figure 1 of North and Vulpe’s review [27]. Tagged (barcoded) strains are pooled and grown together in the presence (treatment) or absence (control) of a compound. After treatment, the molecular barcodes from all strains present in a pooled culture are amplified simultaneously in a single PCR reaction (using a pair of PCR primers that anneal to the common regions flanking the inserted barcodes that are present in all strains). PCR products of barcodes can be hybridized to microarrays containing probe sequences complementary to the barcodes, and the hybridization signal is proportional to the number of cells of that strain in the pooled culture. A fitness score is generated by comparing the hybridization signal of a treated strain to that of a control strain.

HTS technologies have also been used to quantify barcodes in pooled strains. One such method, *Bar-seq*, had improved dynamic range and throughput compared with microarray...
As next-generation sequencing technologies become more affordable, they will predominate. Following barcode quantification, deletion strains whose growth is significantly influenced by the compound treatment can be identified [25, 29]. The deleted genes may be direct targets, or be involved in modifications or pathways that enable the compound’s cytotoxic action [30, 31]. Application of suitable bioinformatics tools, such as Gene Ontology, network analysis and ortholog identification reveal further information about function and pathways in yeast and other organisms [27].

Four different Yeast Knockout mutant collections were created by a consortium of European and North American groups [26] containing homozygous and heterozygous diploid strains corresponding to deletions of 5,916 genes (including 1,159 essential genes) and one haploid strain of each mating type for every non-essential gene (4,757 genes). Each knockout strain is marked by two unique 20-bp barcodes, allowing quantitative and qualitative identification [32].

Screening of yeast deletion mutant collections in many published studies has revealed genes involved in response to toxicants and stressors, including toxic metal ions and pesticides, as reviewed by [15]. We and others have published data on genes involved in resistance to toxic metabolites of arsenic [33], benzene [34], and benzo[a]pyrene [35], formaldehyde [36, 37] as well as mechanisms of toxicity associated with these chemicals. In follow up studies, we validated the roles of human orthologs of some of the yeast genes associated with benzene [38–40] and arsenic [41, 42] toxicity in mammalian cell lines using RNAi-based in vitro knockdown approaches.

2.3 Limitations of screening in yeast

Although yeast functional genomic screening is a powerful tool to identify conserved cellular components required for sensitivity or tolerance to a toxicant treatment, it has certain limitations. First, yeast can tolerate higher level of toxicants than can human cells and thus is not an accurate indicator of toxic doses relevant to humans [33]. Second, information on organ or tissue-specific toxicity and cell-cell signaling is absent. Third, while many genes are conserved between yeast and human, some yeast genes have many human orthologs, making confirmatory experiments challenging. In order to address these issues, similar functional genomic screening technologies are now being developed in higher eukaryotic systems and are discussed in the following sections.

3. Functional genomics in haploid mammalian cells

Mammalian-based screening systems have the potential to generate results that are more directly relevant to toxicity and disease in humans. However, mammals are somewhat tolerant of partial loss of a gene function and inactivation of one gene copy rarely leads to severe changes in phenotype due to the fact that chromosomes are typically diploid in mammals. Therefore, utilization of haploid cells in mammalian screens is necessary. Haploid screening has been established in both human and mouse cells.
3.1 Screening in near-haploid human KBM7 cells

Near-haploid karyotypes have been reported in rare human tumors and leukemias [43] and a heterogeneous (mixed ploidy) cell line (KBM7) was established from the bone marrow of a patient with a near-haploid chronic myeloid leukemia [44]. Although around half of the cells in the initial cultures were near-haploid (apart from disomy of chromosome 8), cells with a diploid or greater DNA content tended to outgrow them with continuous passage, rendering this cell line initially unsuitable for somatic cell genetics. Two years later, this hurdle was overcome when Kotecki et al. reported the derivation of a KBM7 sub-clone (P1-55) that stably remained near-haploid for at least 12 weeks [45].

Carette et al. developed a screening method to generate null alleles for genes on all chromosomes except chromosome 8 using retrovirally-mediated insertional mutagenesis, in the karyotypically stable near-haploid KBM7 sub-clone [46]. Using this approach, they identified host factors essential for infection with influenza and genes involved in cytotoxic response to three bacterial toxins. Using the same approach, Reiling et al. identified a key mediator in the response to tunicamycin, a bacterial protein that induces endoplasm reticulum (ER) stress and the unfolded protein response [47]. The original method developed by Carrette et al. was labor-intensive, and required isolation and expansion of individual clones, followed by DNA extraction, inverse-PCR and Sanger sequencing to map gene-trap insertions. More recently, they incorporated deep sequencing into their screening protocol, facilitating the analysis of millions of mutant alleles in parallel and more accurate assignment of genes to phenotypes [48]. Further, they increased improved genome-wide coverage by increasing the number of cells transfected and the transfection efficiency. Applying this improved approach, they identified 12 human genes important for intoxication by four different cytolethal distending toxins [48] and host factors required for entry of Ebola virus into human cells [49]. The latter was performed in a derivative KBM7 cell line called HAP1, which was generated through a failed attempt to induce pluripotency. HAP1 cells grow adherently, do not express hematopoietic markers and are haploid for all chromosomes including chromosome 8. Importantly, unlike KBM7 cells, HAP1 cells are susceptible to rVSV-GP-EboV, the virus used to test for Ebola-related host factors in the study. Through the inclusion of a unique DNA barcode in each gene-trap vector, Burckstrummer and colleagues subcloned individual gene trap–containing cells, creating a library of isogenic cell lines with mutations in individual genes, enabling more efficient and systematic pooled screens [50]. So far, clones covering 3,396 genes, almost one-third of the expressed genome, have been established [50].

Lehner’s group published two papers illustrating the potential of near-haploid screening to delineate the genes involved in pathogen manipulation of host immune response. First they conducted a forward genetic screen to identify genes required for the function of the Kaposi’s sarcoma herpevirus gene product K5, a ubiquitin ligase that downregulates major histocompatibility complex-1 (MHC-1) and other immunoreceptors [51]. They identified proteolipid 2 (PLP2), a protein of unknown function, as essential for K5 activity and showed that loss of PLP2 traps the viral ligase in the ER, rendering it inactive. Comparison of the plasma proteome of K5-expressing KBM7 cells with and without PLP2, revealed novel targets of downregulation by K5. In a more recent study, Lehner’s group conducted a
forward genetic screen to identify components of the ER-associated degradation (ERAD) pathway targeted by the US11 gene product of human cytomegalovirus, to promote viral immune evasion [52]. They identified TMEM129, a previously uncharacterized membrane protein, as a novel ERAD E3 ubiquitin ligase and central component of a novel ERAD complex that is essential for US11-mediated MHC-1 degradation.

Several groups have utilized near-haploid screening to identify genes and mechanisms of drug toxicity or mechanism of action. Birsoy et al. sought to identify resistance mechanisms to 3-bromopyruvate (3-BrPA), a drug candidate and glycolysis inhibitor [53]. Monocarboxylate transporter 1 (MCT1) was revealed to be a 3-BrPA transporter and key determinant of sensitivity. Further, MCT1 mRNA levels were shown to be predictive of 3-BrPA sensitivity in glycolytic cancer cells, thus representing a potential biomarker of responsive tumors. Reiling et al. performed a near-haploid screen to identify genes involved in apoptosis induced by brefeldin A (BFA), a lead chemotherapeutic compound that is toxic to ER-golgi [54]. A genome-wide haploid genetic screen led to the identification of the small G protein ADP-ribosylation factor 4 (ARF4) that protects against BFA toxicity in a signaling cascade that requires the CREB3 transcription factor. Further, the CREB3-ARF4 pathway was shown to be part of a generalized Golgi stress response. Chen et al. performed a genome-wide genetic screen in haploid human cells to identify genes that confer resistance to severe electron transport chain (ETC) dysfunction when inactivated with a mitochondrial complex III inhibitor, antimycin [55]. Loss of ATP Synthase Mitochondrial F1 Complex Assembly Factor 1 was found to strongly protect KBM7 and other cell types against antimycin-induced, and other forms of ETC dysfunction, and antimycin-induced cell death by facilitating maintenance of mitochondrial membrane potential. In a fourth study, the mechanism by which the genotoxic chemotherapeutic agent YM155 causes DNA damage toxicity was investigated [56]. A requirement for the drug-transporter solute carrier family member 35 F2 (SLC35F2) for YM-155-induced DNA intercalation was discovered and SLC35F2 expression and YM155 sensitivity were correlated across several cancer cell lines.

Though most human haploid screens have involved the selection of mutants resistant to an otherwise lethal agent, thus using cell death or survival as the outcome phenotype, KBM7 screening studies utilizing non-lethality endpoints have been published. Duncan et al. used fluorescence activated-cell sorting to identify genes involved in MHC (major histocompatibility complex) class I antigen presentation, by sorting for mutants that were defective in surface expression of MHC-1 [57]. More recently, Lee et al. used a transcriptional reporter to screen KBM7 cells for constitutive inhibitors of NF-κB and identified previously unknown inhibitors [58].

3.2 Screening in haploid mouse embryonic stem cells

Several groups [59–62] have successfully generated mouse haploid embryonic stem cells (ESC) that can stably grow after multiple passages, be efficiently subcloned, differentiate at similar kinetics as diploid ESCs, and remain in haploid karyotype through initiation of differentiation. Availability of these cells facilitates the application of functional genomics to study the effects of toxicant exposure in normal stem cells and to mimic early-life exposures.
Applying a transposon-based mutagenesis protocol with near genome-wide coverage, Elling et al. identified genes responsible for ricin toxicity [59]. Leeb and Wutz demonstrated and conducted a pilot genetic screen for mismatch repair genes involved in toxicity of 2-amino-6-mercaptopurine [63] and Pettitt et al. identified a gene that mediates olaparib toxicity [64]. These studies demonstrate the potential of functional genetic screening in mouse haploid ESCs and expand the possibility of screening in multiple cell types and developmental pathways [63]. However, mouse ESC are more challenging to culture than human KBM7 cells. Recently, haploid ESC have also been derived from monkey [65] and rat embryos [66].

3.3 Limitations of screening in haploid mammalian cells

One potential disadvantage of KBM7 is that the cells are not completely haploid, having disomy of chromosome 8. Though the KBM7-derivative HAP1, described above, lacks the second copy of chromosome 8, it retains two copies of a fragment of chromosome 15, one of which is fused to chromosome 19 [49]. Very recently, Burckstummer’s group used CRISPR/Cas9-based genome engineering to excise this chromosomal fragment and to derive a truly haploid cell line called eHAP (engineered-HAPloid) [67].

Though this is an important breakthrough, KBM7 and eHAP, as leukemic cell lines from bone marrow, might not reflect the responses of normal cells or other cell types. Carette and colleagues successfully reprogrammed KBM7 cells into induced pluripotent stem cells (iPSC) by infecting the cells with retroviruses carrying four transcription factors [68]. Though the cells retained the BCR-ABL gene fusion, they lost dependency on its signaling for survival and became resistant to imatinib. While KBM7-iPSC have potential as pluripotent cells, they retain other features of the leukemia cells from which they originate. In contrast, haploid mouse ESC are fully haploid and have a largely normal genome. The random mutagenesis approach to generate mutants for haploid screening using gene-trap retroviruses has limitations, including difficulty in reaching genome-wide coverage due to hot and cold spots, and a strong genomic integration bias of retroviruses. Further, it is difficult and time consuming. Targeted, gene-specific approaches such as RNAi and nuclease-mediated editing are required to systematically knockout or knockdown genes in mammalian cells.

4. Functional genomics by RNA interference (RNAi)

4.1 Discovery and application of RNAi

RNAi is an endogenous cellular process, conserved in most eukaryotic species, that involves targeted transcript cleavage and degradation after binding of a sequence-specific short-interfering RNA (siRNA) [69]. This natural process has been exploited as a research tool to target mRNA transcripts by introducing into cells homologous synthetic siRNAs, siRNA precursors such as short-hairpin RNAs (shRNAs), or double stranded RNA (dsRNA) [70–72]. Delivery methods include lipid-based transfection, electroporation and viral transduction with retroviruses and lentiviruses [73]. As some mammalian cell types are resistant to transfection with synthetic siRNAs, they may be alternatively transduced with viruses (e.g. lentivirus) carrying expression vectors (normally plasmid) that encode shRNAs
to express gene-specific siRNAs within the targeted cells, which can achieve stable and highly effective gene suppression in a variety of mammalian cell types [74–76]. Various aspects of RNAi approaches, technical limitations and improvements have been reviewed recently [77].

RNAi has been widely used to study explore the effects of knocking out single genes in functional validation studies. As mentioned in Section 2.2., we used RNAi to knockdown and validate the functional roles of human orthologs of genes identified in yeast screens associated with benzene [38–40] and arsenic [41, 42] in mammalian cell lines. The flexibility of the approach has allowed RNAi to become one of the most powerful tools for genome-wide characterization of gene functions and it is adaptable to HTS. RNAi screens can be conducted in a pooled or arrayed manner (for illustration see Figure 1 in Willingham et al. [78]). In a pooled format, cells are transfected with a siRNA or shRNA library en masse and selected by phenotype following exposure to the agent of interest by microarray or sequencing. The phenotype may be viability or a cell marker selected by fluorescent activated cell sorting. In the arrayed format, cells are transfected with individual siRNAs or shRNAs in separate wells, exposed, and assayed in parallel by a variety of methods, including high-content microscopy or viability screening, or reporter assays [73]. In HCS, cells are labeled with multiple fluorescent markers, which can be measured in multiple channels in a highly automated manner [77].

Vector libraries, such as The RNAi Consortium lentiviral library, which contains shRNAs with unique barcodes targeting 17,200 humans genes [79], have enabled a pooled high-throughput RNAi screening approach called shRNA barcode screening. Individual shRNAs conferring cells with a specific phenotype under the conditions of toxicant treatment can be efficiently identified by PCR amplification of barcodes followed by detection by microarray or sequencing. The magnitude of the effect can be determined by measuring the abundance of identified shRNAs in treated relative to untreated cells. shRNA barcode screening is very effective at identifying genes whose knockdown confers resistance to the test conditions, and has been successfully applied to find genes whose inactivation play a role in breast cancer progression [80]. Most RNAi screens have been based on cell lines or cells with genetic modifications predisposing to certain phenotypes and they have identified genes and gene networks involved in biological processes, drug resistance and pathogen response [72, 81]. For example, in vitro screens have identified genes involved in mitotic arrest and ceramide metabolism as determinants of resistance to chemotherapeutic drugs in different cancer cell types [82] and genes involved in resistance to docetaxel in triple-negative breast cancer cells [83]. RNAi screening has also been applied to understand the functional genomics of both normal and malignant hematopoietic stem cells (HSC) and has uncovered factors regulating differentiation, targets for ex-vivo expansion, tumor suppressor genes associated with hematopoietic cancers and potential therapeutic targets [84, 85]. The GenomeRNAi database (http://www.genomernai.org) curates RNAi phenotype data from the literature for human and Drosophila [86], and contained data on 194 RNAi screens human in its 2014 release (in June, Volume 13).
4.2 Limitations of RNAi screening

Like other functional genomics tools, RNAi screening has limitations. It is limited by the incomplete target gene suppression and confounding OTEs on other mRNA that contain partial sequence homology to the reagent dsRNAs/siRNAs/shRNAs [87–91]. Bioinformatic tools, chemical modification to the seed region, and inclusion of experimental controls can reduce OTEs [72]. Reliable results may require use of more than one dsRNA/siRNA/shRNA targeting different parts of the target mRNA and determination of the silencing efficiency [92]. Functional redundancy exists in protein-coding genes and combinatorial screens targeting two or more genes or a single gene against a genetically sensitized background can be more informative. Aggregate analysis of the data, through pathway or ontology analyses, and integration of the RNAi data with complementary omics data such as transcriptomics can further leverage the power to reveal mechanistic information.

5. Functional genomics by CRISPR-Cas9

5.1 CRISPR-Cas9: Breakthrough genome editing technology

Each of the functional genomic approaches discussed above has strengths and weaknesses but they are all limited in their ability to knock out gene activity on a genome-wide basis in diploid mammalian cells. The recent development of CRISPR-Cas9 technology has provided an effective way of genome editing in a variety of cell types [93–95]. In comparison to RNAi, this RNA-guided technology can generate permanent mutations in the genome, resulting in either a loss or gain of function [95]. CRISPR-Cas9 is a cost-effective, fast, and efficient way to screen many genes. Two other genome editing nuclease technologies preceded CRISPR, zinc-finger nucleases and transcription activator-like effector nucleases (TALENs) [96, 97]. They use protein-DNA interactions for targeting and in contrast with CRISPR, they are more expensive, time-consuming and technically challenging to apply and scale up for screening [98].

CRISPR-Cas9 functions as an adaptive immune system in many bacteria and most archaea by cleaving foreign nucleic acids (e.g. viruses or plasmids) [93, 99]. There are three CRISPR-Cas systems and the simplest, type II CRISPR from Streptococcus pyogenes, utilizes a single Cas9 endonuclease to cleave the DNA and is guided by a 20 nucleotide sequence within an endogenous CRISPR transcribed RNA, and a transacting cRNA. In 2013, George Church’s group demonstrated the potential of the CRISPR-Cas9 system for targeting human cells [100] and four papers were published within a year applying pooled CRISPR screens to target multiple genes within mammalian cells [94, 101–103]. In these approaches, Cas9 is directed by a single-guide RNA (sgRNA) to induce double-strand DNA breaks (DSB) at specific genomic loci. Recognition of cleavage sites and thus target specificity is determined by sgRNA-DNA homology over 20 base pairs and an adjacent protospacer-adjacent motif (PAM), a three nucleotide NGG sequence (where N is any nucleotide) [104]. Web-based tools that facilitate the identification of potential CRISPR target sites in genes of interest are available, e.g., the ZiFiT Targeter software (http://zifit.partners.org/) [105, 106] and the CRISPR Design Tool 51 (http://crispr.mit.edu/) [107]. DSBs are repaired by either the Non-Homologous End Joining (NHEJ) DNA repair pathway or the Homology Directed Repair (HDR) pathway (Figure 1). NHEJ repair yields inserts/deletions (indels) at the DSB site.
leading to frame shifts and/or premature stop codons, resulting in gene knockout. The HDR pathway incorporates a repair template (donor DNA) into the DSB, introducing specific nucleotide changes into a targeted gene [13, 100, 104]. One limitation is that the NHEJ/HDR ratio is difficult to predict in a given cell type.

Use of two gRNAs coupled with Cas9 can efficiently create DNA deletions of up to 10 kb in the presence of a linear homologous repair donor [108]. Interrogation of gene function on a genome-wide scale can be facilitated by large scale oligonucleotide synthesis of guide sequences [109]. In contrast to shRNA libraries, which mediate only gene knockdown, gRNA libraries can be used with Cas9 nuclease to generate libraries of cells with knockout mutations. As reviewed in Sander et al., electroporation, nucleofection and Lipofectamine-mediated transfection have been used to transiently express Cas9 and gRNAs from plasmid DNA in cultured mammalian cells, and lentiviral vectors have been used to constitutively express Cas9 and/or gRNAs in cultured human and mouse cells [98].

Depending on the number of cells and type of cell lines used for screening, a two- or one-vector system to deliver Cas9 and sgRNA is chosen (Figure 2). In the two-vector system, cells are initially transduced with Cas9, after which clones are selected and expanded and subsequently transduced with a sgRNA library. Sabatini and Lander’s group developed a two-vector system, with a library consisting of 73,151 sgRNA plasmids that cover a total of 7,114 human genes with 100 non-targeting controls [101]. The library has been separated into distinct enriched sub libraries for gene targets of known function (e.g. kinases, cell cycle proteins, nuclear proteins, and ribosomal proteins) and genes of unknown function. Using the one-vector system, a large number of cells are transduced with both Cas9 and the sgRNA in a single vector. Such a system was developed by Zhang’s group [94], using a single lentiviral vector to deliver Cas9, a sgRNA, and a puromycin selection marker into target cells. They initially developed a genome-scale CRISPR-Cas9 knockout human library (GeCKOv1) consisting of 64,751 unique guide sequences targeting 18,080 human genes and more recently developed an improved library with 123,411 sgRNAs targeting 19,050 genes (GeCKOv2) [110]. Improvements in the latter library include the ability to target ~1000 additional genes; inclusion of a uniform number (6) of sgRNAs per gene, with three per each of two sublibraries; minimized off-target effects; and sgRNAs that inactivate miRNAs by generating mutations in pre-miRNA hairpin structures. Genome-wide mouse lentiviral sgRNA libraries consisting of 87,897 unique sgRNA plasmids targeting 19,150 protein-coding regions was developed by Koike-Yusa et al. [102] and with 130,209 sgRNAs targeting 20,611 genes by Zhang’s group [110].

5.2 Application of CRISPR-Cas9 screening in functional genomics

Availability of these libraries has enabled genome-wide screening in various mammalian cell lines. In four CRISPR studies using pooled mutant libraries, the first three in human cells and the last in mouse cells, cell death or survival was used as the selection phenotype after exposing the cells to a toxicant and next-generation sequencing was used to identify genes responsible for resistance to each toxicant: nucleotide analog 6-thioguanine and the chemotherapeutic agent etoposide in KBM7 and HL60 cells [101], the protein kinase inhibitor vemurafenib in A375 melanoma cells and HUES62 ESC [94], anthrax and
diphtheria toxin in a focused screen in HeLa cells [103], and Clostridium septicum alpha toxin and 6-thioguanine in mouse ESC [102]. Each study used multiple sgRNAs per gene and statistical tests of enrichment and discovered previously known and novel genes. Genes involved in resistance to vermurafenib correlated well with hits in a previous shRNA screen [94]. Wang and Koike-Yusa identified similar mismatch repair genes involved in 6-thioguanine toxicity in both human and mouse [101, 102].

As well as functional genomic screens, CRISPR-Cas9 is useful for generating genetic models of disease. For example, Heckl et al. used the technology to modify up to 5 genes in a single mouse HSC, leading to clonal outgrowth and myeloid malignancy [111]. Torres et al. generated human cell lines and hematopoietic mesenchymal and stem cells that included chromosomal translocations associated with acute myeloid leukaemia and Ewing’s sarcoma [112]. In another study, human iPSC were modified by a process called iCRISPR, a combination of TALEN and CRISPR-Cas, to generate biallelic knock outs with loss of function, and homozygous knock ins, that could model disease in a stage-specific, inducible manner [113]. CRISPR has also been used to create genetic models of disease in vivo in mice [114, 115]. Very recently, a Cre-dependent Cas9 knock-in mouse was established, enabling the broad application of Cas9 editing in vivo [116]. The utility of this approach was demonstrated by the simultaneous generation of mutations in KRAS, p53, and LKB1 in vivo and dynamic modeling of the resulting lung adenocarcinomas. Rats with conditional alleles in three DNA methylase transferase genes were recently generated using CRISPR-Cas9 [117]. The ability to rapidly generate complex genetic models of disease in vivo enables the examination of gene by environment (GxE) interactions in the development of disease and drug resistance, and for targeted therapy design.

Recently, the ability to repress gene expression by controlling transcription [118–120] and by targeting RNA directly [121] using CRISPR, has been described. Jonathan Weissman’s group at UCSF has developed libraries that enable reversible genome-wide repression by CRISPRi [118] or activation by CRISPRa [122] of gene expression over a wide fold-range [123]. Using a catalytically dead cas 9 protein fused to Krüppel associated box fusion to repress transcription, and a sunCas9-VP16 fusion to activate transcription, they screened for genes involved in response to a chimeric cholera diphtheria toxin (CTx-DTA) in K562 cells and identified genes and pathways involved in pathogen entry, retrotranslocation and toxicity. Previously, haploid mutagenesis revealed the diphthamide biosynthetic pathway (required to generate eEF-2-diphthamide, the target of diphtheria toxin) and the ganglioside biosynthetic pathway (required to produce GM1a, the cell-surface receptor for cholera toxin), as modulating cellular sensitivity to CTx-DTA [124]. CRISPRi provided genetic confirmation that GM1a is the relevant cell-surface receptor for CTx-DTA and revealed the role of additional components in these two pathways and complexity of their responses, enabled by the identification of both sensitizing and protective genes. Additionally, CRISPRi clarified the mechanisms by which the toxin traverses the Golgi network, through the discovery that COG and GARP complexes, which tether late endosomes to the trans-Golgi network or modulate intra-Golgi retrograde transport [125], are critical host factors for CTx-DTA. CRISPRi appeared to approach saturation, evidenced by tight clustering of many of the top hits in protein complexes and pathways. CRISPRa confirmed some of the findings.
of CRISPRi and revealed additional and highly complementary information, e.g. clarifying the roles of enzymes in different branches of the glycosphingolipid biosynthesis pathways in protection against toxicity. These findings highlight some of the advantages of this CRISPR approach over haploid mutagenesis.

A large part of the human transcriptome comprises non-coding RNA, including miRNAs and long non-coding RNA (lncRNA) [126]. Over 56,000 lncRNAs have been discovered to date [127] and roles for some of them in response to xenobiotic exposures have been described [128]. Methodologies to probe the full extent of lncRNA involvement are needed to fully delineate the biological response to xenobiotics. Recent studies have exemplified the power of CRISPR to target non-coding RNAs. Using CRISPRi, Gilbert et al. achieved >80% knockdown for 5 of 6 lncRNAs in K562 [123]. Ho et al. used an approach featuring HR-mediated targeting, NHEJ suppression, and a dual-guide RNA vector to generate knockouts for miRNAs and lncRNAs [129]. As mentioned above, the GeCKOv2 library developed by Feng Zhang’s group contains sgRNAs that target human miRNAs (n=1,864) [110] by directing mutations to the pre-miRNA hairpin structure [130].

5.3 Limitations of CRISPR-Cas9 screening

CRISPR-Cas9 has limitations that need to be addressed before it can be deployed, particularly in the therapeutic realm, on a large scale. Though CRISPR-Cas9 has a high validation rate, OTEs do occur partly due to the small seed region guiding specificity, albeit at a lower rate than with RNAi [131]. Several strategies to reduce OTEs have been developed. For example, a paired nickase or dual nickase strategy, in which adjacent offset nicks are generated at the target site using two gRNAs and two co-dependent Cas9 nickase monomers, has been described [98, 132–134]. Joung’s group reduced OTEs substantially by simply using sgRNAs that were truncated (tru-gRNAs) by 2–3 nucleotides at the 5’ end of their complementarity regions [135]. Another approach to minimize off-target effects is to predict them in silico, e.g. the bioinformatics tool COSMID (CRISPR Off-target Sites with Mismatches, Insertions, and Deletions) searches genomes for potential off-target sites (http://crispr.bme.gatech.edu), helping to inform the design of CRISPR-Cas systems with minimal off-target effects [136]. In a recent paper, Yang et al. used whole-genome sequencing to assess target specificity in human-induced pluripotent stem cells (hiPSC) and found that single nucleotide variants can create OTEs, with a likelihood of 1.5–8.5%, depending on the genome and site-selection method and thus has important implications for subject-specific design [127]. Joung’s group developed an approach called genome-wide, unbiased identification of DSBs enabled by sequencing (GUIDE-seq), that relies on capture of double-stranded oligodeoxynucleotides into DSBs. The method identified many OTEs that were not detectable by existing computational methods and confirmed that tru-gRNAs exhibit reduced numbers of OTEs [137].

Another limitation is that the range of currently available target sites is restricted by the need for a PAM sequence matching the form NGG; use of alternative PAM sequences is being explored but requires further validation [98]. Use of gRNA in conjunction with Cas9 from other species may also expand target options.
Strategies to reduce OTEs, to increase the target range and specificity, and shift the balance away from NHEJ-mediated indel mutations and toward HDR-driven alterations are priorities. A final limitation is that some cell lines are more challenging to transfect/infect, and may limit the technology to less relevant cell lines.

Despite these challenges, the studies to date demonstrate the potential of the CRISPR-Cas9 system for conducting large-scale genome-wide screens in mammalian cells. This system offers several powerful features such as gene inactivation at the genomic DNA level and lower off-target effects than RNAi, which will be further reduced by ongoing improvements to the technology.

6. Potential of CRISPR-Cas9 for toxicity screening

Each of the screening methodologies described in this paper has advantages and disadvantages as discussed above. The CRISPR-Cas9 screening approach, which is evolving rapidly, overcomes many of the limitations of previous approaches. It can be used to repress and activate both coding and non-coding genes of interest, with fewer OTEs than RNAi. It is compatible with any human cell type that can be transfected and thus can identify genes that are more directly relevant to normal human cells compared with yeast and KBM7 haploid screening. Libraries are commercially available that can target a large and growing number of human genes. Therefore, CRISPR-Cas9 is a powerful functional genomic screening approach that will provide unprecedented mechanistic insight in the field of modern toxicology.

First, as described earlier, CRISPR-Cas9 has been used to identify genes involved in the response to chemical and microbial toxicants in several cell types [94, 101–103]. Thus, it could readily be extended to the systematic screening of large numbers of environmental chemicals in a variety of human cell lines. Second, as well as the knockout of gene function, CRISPRi and CRISPRa facilitate the repression and activation of gene expression with near-saturation [123]. Thus, it offers the potential to more fully characterize AOPs and AOP networks than yeast or human haploid screening. Third, as a flexible gene-editing tool, CRISPR-Cas9, as well as allowing the examination of the effects of a large number of gene knock-outs simultaneously, enables the precise targeting of individual genes and gene regions identified by CRISPR, or other screening methods, for further validation. Fourth, as well as screening genes, CRISPR-Cas9 can be used to precisely generate human cell and animal models of development and disease [111–117] that can increase the efficiency of preclinical toxicity testing and be used to discover or validate GxE interactions at the level of individual genotypes or haplotypes. Fifth, CRISPR-Cas9 can be used to target non-coding RNAs [110, 123, 129, 130], which represent a large proportion of the human transcriptome and potentially play important roles in response to toxicants.

Ongoing improvements in target range and specificity, automation, next generation sequencing, and data handling and analysis, will increase the throughput of screening techniques. High-throughput screening of chemicals of concern or unknown toxicity will elucidate common and unique mechanisms and pathways of toxicity. The time is right for the establishment of efforts to systematically evaluate chemicals of concern by CRISPR-
Cas9 screening in human cells. Such an approach could be designed to address some of the major challenges of risk assessment today [1], including the identification of individual variation in susceptibility at the level of genes and toxicity pathways, as well as understanding the effect of environmental levels of exposure, co-exposures, tissue-specific effects (use of different cell types), and early-life effects (use of hESC and differentiated cells).

Acknowledgments

We acknowledge Professor Chris Vulpe for his long-term collaboration with us on screening of environmental chemicals using functional genomics in the yeast. Dr. Hua Shen is a postdoctoral trainee in the Superfund Research Program at the University of California Berkeley. This work was supported by the U.S. National Institute of Environmental Health Sciences Superfund Research Program P42 ES004705 (MTS and LZ) and by the U.S. National Institute of Health grant R01ES017452 (LZ).

Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>3-BrPA</td>
<td>3 bromopyruvate</td>
</tr>
<tr>
<td>AOP</td>
<td>adverse outcome pathway</td>
</tr>
<tr>
<td>ARF4</td>
<td>ADP-ribosylation factor 4</td>
</tr>
<tr>
<td>BFA</td>
<td>brefeldin A</td>
</tr>
<tr>
<td>BT</td>
<td>1,2,4-benzenetriol</td>
</tr>
<tr>
<td>CML</td>
<td>chronic myeloid leukemia</td>
</tr>
<tr>
<td>COSMID</td>
<td>CRISPR off-target sites with mismatches, insertions, and deletions</td>
</tr>
<tr>
<td>CRISPR-Cas9</td>
<td>clustered regularly interspaced short palindrome repeats-associated nuclease-Cas9</td>
</tr>
<tr>
<td>CTx-DTA</td>
<td>chimeric diphtheria toxin</td>
</tr>
<tr>
<td>DSB</td>
<td>double-strand break</td>
</tr>
<tr>
<td>dsRNA</td>
<td>double-stranded RNA</td>
</tr>
<tr>
<td>eHAP</td>
<td>engineered HAPloid</td>
</tr>
<tr>
<td>ETC</td>
<td>electron transport chain</td>
</tr>
<tr>
<td>ER</td>
<td>endoplasmic reticulum</td>
</tr>
<tr>
<td>ERAD</td>
<td>ER-associated degradation product</td>
</tr>
<tr>
<td>ESC</td>
<td>embryonic stem cell</td>
</tr>
<tr>
<td>GeCKo</td>
<td>genome-scale CRISPR-Cas9 knockout</td>
</tr>
<tr>
<td>GWAS</td>
<td>genome-wide association studies</td>
</tr>
<tr>
<td>GUIDE-seq</td>
<td>genome-wide unbiased identification of DSBs enabled by sequencing</td>
</tr>
<tr>
<td>HCS</td>
<td>high-content screening</td>
</tr>
<tr>
<td>HDR</td>
<td>homology-directed repair</td>
</tr>
</tbody>
</table>
HQ hydroquinone
HTS high-throughput screening
iPSC induced pluripotent stem cell
lncRNA long non-coding RNA
MHC major histocompatibility protein 1
MCT1 monocarboxylate transporter 1
MOA mode of action
NGS next-generation sequencing
NHEJ non-homologous end-joining repair
ORF open reading frame
OTE off-target effects
PAM protospacer-adjacent motif
PCR polymerase chain reaction
PDA parallel deletion analysis
PLP2 proteolipid 2
RNAi RNA interference
sgRNA single-guide RNA
tru-gRNA truncated sgRNA
shRNA short hairpin RNA
siRNA short-interfering RNA
SLC35F2 solute carrier family member 35 F2
TALEN transcription activator-like effector nuclease

References


51. Shen et al. Mutat Res Rev Mutat Res. Author manuscript; available in PMC 2016 April 01.


Figure 1.
CRISPR-Cas9 DSB repair pathway and gene editing [13]. Reproduced with permission from *Nature Protocols.*
Figure 2.
Lentiviral vector delivery system for Cas9 and sgRNA [138].
Table 1

Examples of recent functional genomic screening approaches used in toxicity studies

<table>
<thead>
<tr>
<th>Methodology</th>
<th>Advantages</th>
<th>Limitations</th>
<th>Genome Coverage</th>
<th>Cell Type</th>
<th>Organism</th>
<th>Toxic Agent</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCR-based gene disruption</td>
<td>Yeast genome well annotated; many genes conserved between yeast and human; quantitative; high-throughput.</td>
<td>Not all yeast genes have human homology; yeast tolerates higher levels of toxicants than human cells do.</td>
<td>All nonessential genes</td>
<td>Yeast</td>
<td>S. cerevisiae</td>
<td>Arsenic</td>
<td>[33]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Insertional mutagenesis</td>
<td>Extends yeast approach to human cells</td>
<td>Not haploid on chromosome 8; strong genomic integration bias limits true coverage; leukemic cells.</td>
<td>All human genes except those on chromosome 8</td>
<td>KBM7 haploid CML cell line</td>
<td>H. sapiens</td>
<td>3-bromopyruvate</td>
<td>[53]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transposon mutagenesis</td>
<td>Extends yeast approach to mouse cells</td>
<td>Strong genomic integration bias</td>
<td>Genome wide</td>
<td>Mouse haploid ESC</td>
<td>M. musculus</td>
<td>2-amino-6-mercaptopurine</td>
<td>[63]</td>
</tr>
<tr>
<td>Methodology</td>
<td>Advantages</td>
<td>Limitations</td>
<td>Genome Coverage</td>
<td>Cell Type</td>
<td>Organism</td>
<td>Toxic Agent</td>
<td>Reference</td>
</tr>
<tr>
<td>-------------</td>
<td>------------</td>
<td>-------------</td>
<td>-----------------</td>
<td>-----------</td>
<td>----------</td>
<td>-------------</td>
<td>-----------</td>
</tr>
<tr>
<td>RNAi</td>
<td>Targets RNA; applicable to many cell types; libraries targeting human genes available.</td>
<td>Incomplete knock down; off target effects (OTE); unsuitable for lethal phenotypes.</td>
<td>Genome-wide at RNA level</td>
<td>Human cancer cell lines</td>
<td>H. sapiens</td>
<td>Anti-cancer therapeutics</td>
<td>[81–83]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Insect</td>
<td>D. melanogaster</td>
<td>Cell cycle inhibitors</td>
<td>[141]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Worm</td>
<td>C. elegans</td>
<td>Paraquat</td>
<td>[142]</td>
</tr>
<tr>
<td>CRISPR-Cas9</td>
<td>Precise targeting of DNA and RNA; applicable to many cell types; permanent mutations; complete knockout; libraries targeting human genes available.</td>
<td>Only cleave DNA at loci with a suitable proto-spacer adjacent motif (PAM) sequence; OTE</td>
<td>Genome wide</td>
<td>Cancer cell lines and ESC</td>
<td>H. sapiens</td>
<td>6-thioguanine</td>
<td>[101]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Diphtheria toxin</td>
<td>[123]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Etoposide</td>
<td>[101]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Vemurafenib</td>
<td>[94]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6-thioguanine</td>
<td>[102]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>M. musculus</td>
<td>Alpha-toxin</td>
<td>[102]</td>
</tr>
</tbody>
</table>

CML, chronic myeloid leukemia; HQ, hydroquinone; BT, benzenetriol; CAT, catecholamine; OTE, off-target effects; PAM, proto-spacer adjacent motif; ESC, embryonic stem cells