Chapter 7
Target Discovery for Precision Medicine Using High-Throughput Genome Engineering

Xinyi Guo, Poonam Chitale, and Neville E. Sanjana

Abstract Over the past few years, programmable RNA-guided nucleases such as the CRISPR/Cas9 system have ushered in a new era of precision genome editing in diverse model systems and in human cells. Functional screens using large libraries of RNA guides can interrogate a large hypothesis space to pinpoint particular genes and genetic elements involved in fundamental biological processes and disease-relevant phenotypes. Here, we review recent high-throughput CRISPR screens (e.g. loss-of-function, gain-of-function, and targeting noncoding elements) and highlight their potential for uncovering novel therapeutic targets, such as those involved in cancer resistance to small molecular drugs and immunotherapies, tumor evolution, infectious disease, inborn genetic disorders, and other therapeutic challenges.

Keywords Genome engineering • Pooled CRISPR screens • Functional genomics • Cancer • Drug resistance • Infectious disease • Metabolism • Target identification

7.1 Introduction

The recent development of RNA-guided CRISPR nucleases for genome editing has created new opportunities for understanding the genetic basis of disease. With the development of pooled screens utilizing RNA-programmable nucleases, thousands of genes can be interrogated simultaneously to test many genetic hypotheses in parallel. Beyond their initial application for loss-of-function screening, pooled CRISPR screens have also been adapted for gene overexpression, repression, and enhancer region modulation. Here, we first present an overview of pooled screen workflows and how different CRISPR effectors can be harnessed to activate, repress, or knockout genes in different disease models (Fig. 7.1a). We also survey
applications of CRISPR screens in cancer, infectious diseases and inborn genetic disorders (Fig. 7.1b and Table 7.1). We highlight how these screens have been used for target discovery and potential therapeutic developments from identified target genes/genetic elements.

7.2 Technologies for CRISPR Screens

7.2.1 From Gene Editing to Pooled Screens

Programmable nucleases, such as the clustered regularly interspaced short palindromic repeat-associated nuclease Cas9 (CRISPR/Cas9) have ushered in a new era of precise genome manipulation. For targeted modification in mammalian cells, it is necessary to express both the Cas9 nuclease and a single-guide RNA (sgRNA) [1–3]. The sgRNA contains a 20 nt sequence complementary to the target
<table>
<thead>
<tr>
<th>Disease category</th>
<th>Disease</th>
<th>Screening phenotype</th>
<th>Therapeutic implications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cancer</td>
<td>Melanoma</td>
<td>Resistance to chemotherapy drug (a BRAF protein kinase inhibitor vemurafenib [6, 24, 32]); lethal/anti-proliferative phenotype [43, 85]</td>
<td>Screened human coding genes for loss-of-function (LoF) and gain-of-function (GoF) mutations that confers vemurafenib resistance in BRAF mutant melanoma cancer cell line, validated known resistance genes, and identified novel targets (LoF genes such as NF2 and CUL3 [6] and GoF genes such as ITGB3 and P2RY8 [22]). Further study expanded the screening region to search for adjacent noncoding regulatory elements that mediate the expression of vemurafenib resistance genes (e.g. CUL3) and identified a list of previous uncharacterized regulatory elements and mechanisms [32]. Screened human coding genes and identified melanoma-specific cancer dependencies [43, 85]</td>
</tr>
<tr>
<td>Myeloid leukemia</td>
<td>Resistance to chemotherapy drugs such as etoposide [7], cytosine arabinoside (Ara-C) [48] and ATR kinase inhibitor [49]; protein stability reporter assay [86]; lethal/anti-proliferative phenotype [87–89]</td>
<td>Screened human coding genes for LoF mutations that confer drug resistance in myeloid leukemia cell lines, validated known resistance genes (e.g. DCK gene for Ara-C resistance in acute myeloid leukemia, (AML)) [48], and identified new resistance genes (e.g. CDK6 for etoposide resistance in chronic myeloid leukemia, (CML) [7], SLC29A in Ara-C resistance [48] and CDC25A in ATR inhibitor resistance in AML [49]), provided therapeutic insights into combinatorial drug actions based on patient’s genotypes (e.g. prednisolone for DCK negative cases [48] and WEE1 inhibitors for ATR inhibitor resistant cases [49]). Another study screened human coding genes for regulators of resistance factors and identified new genes regulating CDC25A protein stability in ATR inhibitor resistance [49]. AML-specific genetic vulnerabilities have been identified with a genome-wide screen of human protein-coding genes [87–89] and saturation mutagenesis of 192 chromatin regulatory domains in the mouse genome [90]. These studies have expanded the list of druggable genetic dependencies in AML, identified new regulatory microRNAs required for proliferation [89], and highlighted new candidates (such as KAT2A [87], ENL [88]) for pharmacological inhibition.</td>
<td></td>
</tr>
<tr>
<td>Disease category</td>
<td>Disease</td>
<td>Screening phenotype</td>
<td>Therapeutic implications</td>
</tr>
<tr>
<td>------------------</td>
<td>---------</td>
<td>---------------------</td>
<td>-------------------------</td>
</tr>
<tr>
<td>Non-Hodgkin’s lymphoma</td>
<td>Apilimod</td>
<td>cytotoxicity</td>
<td>[91]; induction of lymphoma [92]</td>
</tr>
<tr>
<td></td>
<td>Lethal/anti-proliferation phenotype</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Hepatocellular carcinoma</td>
<td>Deletion-based screen of human long noncoding RNA (lncRNAs) in HuH-7; PiggyBac-based transposon screen</td>
<td>[46, 93]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>This work suggests meta-transcriptional regulator HMGA2 (part of the NF1 pathway) as a potential inhibitory target to treat liver cancer [93].</td>
</tr>
<tr>
<td>Lung adenocarcinoma</td>
<td>Genome-wide in vivo pooled competition assay for pro-growth and pro-metastasis LoF mutations.</td>
<td>Induction of primary tumor and lung-metastatic phenotype [94]; lethality-anti-proliferation phenotype [94];</td>
<td></td>
</tr>
<tr>
<td>Neuroblastoma</td>
<td>Screened human coding genes for LRCC cancer dependencies in NRAS-mutant neuroblastoma cell line; validated driver mutations and their downstream kinases in the screen [94]; discovered putative dependencies including <strong>TBK1</strong> gene [94].</td>
<td>Lethal/anti-proliferation phenotype</td>
<td></td>
</tr>
<tr>
<td>Disease</td>
<td>Phenotype</td>
<td>Therapeutic implications</td>
<td></td>
</tr>
<tr>
<td>--------------------</td>
<td>------------------------------------</td>
<td>--------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Sarcoma</td>
<td>Lethal/anti-proliferation phenotype</td>
<td>Established a patient-derived sarcoma cell line, and screened a selective set of druggable human genes combining CRISPR, RNAi, and small-molecule screening approaches. The study identified CDK4 and XPO1 as potential therapeutic targets [44].</td>
<td></td>
</tr>
<tr>
<td></td>
<td>myogenic differentiation [95]</td>
<td>Suggested HDAC3 as a potential therapeutic target in myogenic differentiation therapy (where malignant cells are encouraged to differentiate into non-proliferative cells) [95].</td>
<td></td>
</tr>
<tr>
<td>Breast cancer</td>
<td>Oncogene-induced senescence (anti-proliferation phenotype) [96]</td>
<td>Screened 90% of p53-bound enhancers and 60% of ERα-bound enhancers for functional elements. The enriched functional elements for the senescence phenotype (for p53) or growth phenotype (for ERα) were enhancers near cell-cycle genes (CDKN1A and CCND1) [96].</td>
<td></td>
</tr>
<tr>
<td>Ovarian cancer</td>
<td>Lethal/anti-proliferation phenotype</td>
<td>Screened 50 epigenetic regulators with a paired-gene targeting sgRNA library to study cooperative regulation and possibly co-regulate cancer and stromal cell invasion. Identified a KDM4C inhibitor and BRD4 inhibitor to reduce ovarian cancer cell proliferation [34].</td>
<td></td>
</tr>
<tr>
<td>Ovarian cancer</td>
<td>Lethal/anti-proliferation phenotype</td>
<td>Screened human coding genes in RNF43-mutant pancreatic ductal adenocarcinoma cell line, for cancer-dependent specificity in proliferation, identified and functionally validated the FZD5 receptor as a target for potential targeted therapy using antibody [97].</td>
<td></td>
</tr>
<tr>
<td>Pancreatic cancer</td>
<td>Lethal/anti-proliferation phenotype</td>
<td>Screened human coding genes in KRAS-mutant colorectal cancer cell lines, discovered distinct vulnerabilities between common KRAS-mutant colorectal cancer cell lines: DLD1, appears to be dependent on EGFR signaling (despite the cells being KRAS-mutant), whereas HCT16 relies on ETC complex I function and can be inhibited selectively by methotrexate [43].</td>
<td></td>
</tr>
<tr>
<td>Colon cancer</td>
<td>Lethal/anti-proliferation phenotype</td>
<td>Screened mouse genome for resistance to α-toxin or 6-thioguanine; identified 27 previously characterized and 4 novel genes involved in resistance [98].</td>
<td></td>
</tr>
<tr>
<td>Infections disease</td>
<td>Clostridium septicum infection</td>
<td>Resistance to either clostridium septicum alpha (α)-toxin or 6-thioguanine. Identified 27 previously characterized and 4 novel genes involved in resistance [98].</td>
<td></td>
</tr>
</tbody>
</table>

(continued)
<table>
<thead>
<tr>
<th>Disease category</th>
<th>Disease</th>
<th>Screening phenotype</th>
<th>Therapeutic implications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Synthetic bacterial toxins</td>
<td>Sensitivity to diphtheria and chimeric anthrax toxins [57]; sensitivity to a cholera-diphtheria toxin [16]</td>
<td>Screened 291 genes for host factors essential for anthrax and diphtheria intoxication; identified four genes (PLXNA1, FZD10, PECR and CD81) as candidates in protective agent-mediated anthrax toxicity [57]. Screened human coding genes for sensitizing and resistant host factors that regulate cellular response to cholera-diphtheria intoxication [16].</td>
<td></td>
</tr>
<tr>
<td>Gram-negative bacterial infections</td>
<td>Inflammatory cytokine tumor necrosis factor (Tnf) positive after lipopolysaccharide (LPS) stimulation</td>
<td>Screened mice genome for induction of Tnf after LPS stimulation, identified novel genes that regulate TLR4 signaling pathway in response to LPS [99].</td>
<td></td>
</tr>
<tr>
<td>Staphylococcus aureus infection</td>
<td>Sensitivity to α-hemolysin toxin (αHL)</td>
<td>Screened human coding genes and identified 10 host targets required for αHL susceptibility; validated proteins (such as SYS1, ARFRP1, and TSPAN14) that regulate host receptors [100].</td>
<td></td>
</tr>
<tr>
<td>Vibrio parahaemolyticus infection</td>
<td>Sensitivity of type III secretion system (T3SSs)-dependent cytotoxicity</td>
<td>Screened human coding genes for distinct host factors facilitating T3SSs-dependent cytotoxicity, discovered processes underlying host-pathogen interactions [59].</td>
<td></td>
</tr>
<tr>
<td>West Nile virus (WNV) infection</td>
<td>Resistance to West Nile virus (WNV) infection [63, 64]</td>
<td>Screened human coding genes for host factors required for West Nile virus infectivity; identified genes including 7 strongly protective genes (EMC2, EMC3, SEL1L, DERL2, UBE2G2, UBE2J1, and HRD1) in the ER-associated protein degradation (ERAD) pathway as potential therapeutic targets to prevent WNV induced cell death [63]. Another CRISPRn screen identified and specifically detailed the role of SPCS1 in modification and secretion of flaviviral particles, and suggested that inhibition of SPCS1 might reduce viral replication [64].</td>
<td></td>
</tr>
</tbody>
</table>

Table 7.1 (continued)
<table>
<thead>
<tr>
<th>Disease category</th>
<th>Disease Screening phenotype</th>
<th>Therapeutic implications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dengue virus (DENV) and hepatitis C virus (HCV) infection</td>
<td>Resistance to dengue virus (DENV) and hepatitis C virus (HCV) infection</td>
<td>Screened human coding genes for host factors required for DENV and HCV infectivity; identified oligosaccharyltransferase complex as essential elements for DENV replication and discovered role of intracellular flavin adenine dinucleotide during HCV replication which suggests new host targets for antiviral drugs [65].</td>
</tr>
<tr>
<td>Human immunodeficiency virus (HIV)</td>
<td>Lethal phenotype</td>
<td>Screened human coding genes and identified five HIV host cell factors: co-receptors CD4 and CCR5, TPST2, SLC35B2, and ALCAM, which suggest new cellular pathways for antiviral therapy [66].</td>
</tr>
<tr>
<td>Chronic viral infection</td>
<td>Altered PD-1 expression profile</td>
<td>Screened ~23.8 kb enhancer and eight additional regulatory regions adjacent to Pdcd1 locus in murine T cells; discovered that exhausted CD8+ T cells have a unique enhancer and transcription factor binding landscape which suggests an option to edit exhaustion-specific enhancers for engineered T-cell therapy [52].</td>
</tr>
<tr>
<td>Apicomplexan parasites</td>
<td>Lethal phenotype</td>
<td>Screened T. gondii parasite genome for factors that facilitate infection; identified ~200 new fitness genes and investigated critical factors (such as CLAMP protein) involved in host cell infection [70].</td>
</tr>
<tr>
<td>Inborn genetic disorders</td>
<td>Fetal hemoglobin (HbF) enrichment [31, 33]</td>
<td>Screened BCL11A composite enhancer DNase I hypersensitive sites in human hematopoietic stem and progenitor cells; identified a conserved GATA1 motif as the essential element of BCL11A enhancer for human erythroid BCL11A expression and HbF repression, which suggests a potential therapeutic genome editing site for β-hemoglobin disorders [31]. Screened HBS1L-MYB intergenic region (whose variants modulate HbF level) for regulatory elements that control MYB expression; identified putative regulatory elements that control MYB expression [33].</td>
</tr>
<tr>
<td>Mitochondrial diseases</td>
<td>Lethal phenotype [74, 75, 77]</td>
<td>Screened human coding genes for protective factors during mitochondrial respiratory chain (RC) inhibition; identified von Hippel-Lindau (VHL) factor and proposed a hypoxia treatment for mitochondrial disease [74]. Screened human genome for novel genes essential for oxidative phosphorylation; identified 191 hits and discovered NGRN, WBSCR16, RPUSD3, RPUSD4, TRUB2, and FASTKD2 that form a mitochondrial 16S rRNA regulatory module [75]. Screened human genome for LoF mutations that can rescue Mitochondrial complex I-impaired cytoplasmic hybrid cells in conditions that require cellular oxidative phosphorylation; identified BRD4 whose loss enhances oxidative phosphorylation activity, suggesting it as a promising target to overcome mitochondrial defects [77].</td>
</tr>
</tbody>
</table>
genomic region, and a part of a palindromic repeat that forms the secondary structure for Cas9 docking [4]. Directed by sgRNA, Cas9 nuclease identifies a target genomic region and introduces a double-stranded break (DSB). Chromosomal DSBs are typically repaired through cellular repair mechanisms such as homologous recombination (HDR) or non-homologous end joining (NHEJ). In the NHEJ repair pathway, the Ku DNA-binding heterodimer first binds to the DNA terminus to initiate end processing and recruits enzymes such as Artemis-DNA-PKcs to trim the incompatible ends, polymerases to fill the gaps, and ligases (XRCC4-DNA-ligase-IV complex) to seal the nick [5]. Cellular repair mechanisms such as NHEJ often create deletions or insertions (indel mutations) at the DSB site. If Cas9 targets a coding exon, indel mutations can result in a frameshift mutation and a premature stop codon, thereby knocking out the target gene expression. If Cas9 targets an intron, enhancer, or other noncoding region, mutagenesis can disrupt functional elements such as transcription factor binding motifs or chromatin anchoring sites, which can alter regulation of gene expression.

CRISPR forward genetic screens take advantage of the same genome editing machinery to pair many different genetic changes with a phenotypic assay [6, 7]. Specifically, the screen quantifies which genetic manipulations are enriched or depleted in a disease-relevant phenotype. The workflow for CRISPR screens can be summarized in five steps: (1) choose genomic regions or genes of interest and design a sgRNA library to target these elements, (2) generate cell populations with various genetic perturbations introduced through this sgRNA library, (3) select a biological phenotype of interest, (4) trace back from the selected phenotype to its associated gene/genomic targets, and (5) confirm the function of the identified targets through additional validation studies [8].

CRISPR screens can be performed in either an arrayed or pooled format. In an arrayed CRISPR screen, each well receives one sgRNA delivered into all cells. In comparison, a pooled CRISPR screen can perturb thousands of genes simultaneously—with each cell in the pool receiving one genetic perturbation. This is most often achieved via lentiviral delivery of the CRISPR library to a large cell pool. Each construct in the pooled lentiviral library contains a unique sgRNA. To ensure that each cell only receives a single CRISPR construct, the viral titer is adjusted such that the multiplicity of infection is less than 1 (i.e. fewer viral particles than cells). Successful genomic integration of the virion results in expression of the sgRNA in a Cas9-expressing cell line. Alternatively, both sgRNA and Cas9 nuclease can be packed into the same virion to infect wild type cell lines. To remove non-transduced cells, the construct also includes a selectable marker such as drug resistance or fluorescence. After lentiviral integration, the unique 20 nt sgRNA guide sequence serves as a barcode for the construct. This barcode is used to measure enrichment or depletion of the specific sgRNA after phenotypic selection. Significant enrichment or depletion of a sgRNA barcode suggests functional association between the sgRNA target locus and the phenotype of interest. To reduce false-positive hits, genes/genome target regions should be validated with newly-designed
sgRNAs that are not in the original library. Validation of individual sgRNAs should also include analysis of indels (e.g. Surveyor/T7E1, sequencing, etc.) and/or gene expression changes (e.g. qPCR, quantitative protein blotting, etc.). After initial hit validation, further in-depth studies may involve genetically-engineered mouse models, perturbations of related genes in the same pathway, and validation across a panel of cell lines to examine the effects of genetic background.

7.2.2 Types of CRISPR Screens

CRISPR screens to date have mainly focused on applying CRISPR nuclease (CRISPRn) Cas9 to identify loss-of-function mutations in protein-coding genes associated with disease traits. In addition to their use as a targeted nuclease, CRISPR systems have also been deployed as a general DNA-targeting platform to bring new effector domains to specific regions of the genome [9–11]. Beyond Cas9, there are also exciting possibilities for applying other DNA and RNA targeting CRISPR systems to take advantage of the metagenomic diversity of CRISPR systems [12]. These different CRISPR systems and effector domains can greatly diversify the genetic manipulations available for screening gene loci and noncoding regions.

There is a variety of effector fusions that have been developed to activate or repress gene expression. Gene repression via effector domains is distinct from nuclease-based gene loss-of-function. Cas9 nuclease targeting typically results in loss-of-function due to formation of indel mutations in coding exons and nonsense-mediated decay of mRNA transcripts. In contrast, CRISPR interference (CRISPRi) screens use a deactivated Cas9 (dCas9) fused to a Krüppel-associated box domain (KRAB) repressor [13]. Deactivated Cas9 (via alanine mutagenesis of a catalytic residue in the nuclease domain) retains the ability to form Cas9-sgRNA complexes that bind target sites [14]. The KRAB repressor is one of the most commonly used effectors for gene repression. Once at the target site, KRAB recruits nuclear proteins to form a heterochromatin complex that can facilitate histone methylation and deacetylation [15]. CRISPRi screens using dCas9-KRAB have been applied to study protective factors in cellular toxin-resistance [16] and identify regulatory elements in the vicinity of oncogenes such as \textit{GATA1} and \textit{MYC} [17]. For upregulating gene expression, there are three major types of dCas9-based gene-activating approaches (CRISPRa): tethering dCas9 directly with one or multiple activators (dCas9-VP64 [18, 19], dCas9-VPR [20], dCas9-P300 [21], and dCas9-VP160 [22]); engineering a polypeptide scaffold to dCas9 for tagging multiple activator copies (Suntag [23]); modifying sgRNA scaffold hairpin region to recruit activators (SAM [24] and others [25]). A recent comparison of dCas9 activators found that activators with multiple, distinct activation domains (dCas9-VPR, SAM and Suntag) were capable of higher and more robust gene activation compared to effectors with
In addition to gene activation and repression, other effectors have been incorporated into CRISPR systems to manipulate DNA methylation, histone acetylation and base editing. DNA methylation is catalyzed by DNA methyltransferases (Dnmt) and typically results in gene silencing [26] whereas DNA demethylation is facilitated by ten-eleven translocation (TET) dioxygenases and can result in gene activation [27]. Catalysts of DNA methylation and demethylation can be fused with dCas9, such as dCas9-Dnmt3a and dCas9-Tet1 respectively, and have been used to precisely edit CpG methylation [27]. Recent studies have shown that DNA methylation correlates with certain neuropsychiatric disorders such as schizophrenia, Rett syndrome, and immunodeficiency-centromeric instability (ICF) syndrome [26]. CRISPR screen effectors dCas9-Dnmt3a or dCas9-Tet1 could be used to identify regions of the genome that harbor control elements sensitive to changes in methylation. In addition to DNA methylation, post-translational modifications to histone tails can also modulate gene expression. Fusing the catalytic unit of acetyltransferase to dCas9 can robustly activate gene expression by catalyzing acetylation of histone H3 lysine 27 at enhancer/promoter sites [21]. Additionally, the base pair editing tool dCas9-cytidine deaminase fusion protein has been used for making C to T (or G to A) point mutations [28]. Another point mutation generator system: “CRISPR-X” used dCas9 and a modified sgRNA with two MS2 hairpins to recruit a cytidine deaminase [29]. These systems can act as re-purposed CRISPR screens to provide alternatives to the kinds of mutations that result from CRISPRn-driven NHEJ.

Recently, pooled screens that pair CRISPR nucleases with multiple guides have been used to analyze multi-gene interactions and larger deletions. To study noncoding elements such as long noncoding RNAs (lncRNAs) or super-enhancers, pairs of sgRNAs can create deletions that span the beginning and the end of larger genomic regions. A deletion screen targeting multiple long noncoding RNAs successfully demonstrated targeted genomic deletions to pinpoint regulatory lncRNAs associated with liver cancer [30]. For higher-resolution tiling in the noncoding region, single sgRNA saturation mutagenesis has been particularly helpful in identifying functional elements such as transcription binding motifs [31]. A saturating-mutagenesis screen targeting ~700 kb region surrounding drug resistance genes has uncovered regulatory elements in a melanoma model [32]. Another study utilized a saturating-mutagenesis library to examine ~300 kb region in HBS1L-MYB intergenic region and identified putative enhancer elements that regulates MYB expression, which in turn regulates fetal hemoglobin levels [33]. Multi-guide screens have also been used to search for loss-of-function gene interactions or cooperative regulatory networks [34].

In addition to different effectors, CRISPR screens can benefit from the abundance and diversity of CRIPSR-based DNA-targeting/gene editing systems found in different microbial species. Recent work on the CRISPR effector Cpf1, which recognizes T-rich PAMs [35, 36], suggests a new screening option for targeting T-rich,
NGG-poor regions. Since Cpf1 processes its own repeat array through its ribonuclease activity, it may be easier to multiplex guide RNAs [37] for examining cooperative regulation and deletions. To further expand the screening target from genome to transcriptome, the recently discovered RNA editing Cas9-C2c2 [38, 39] could be deployed to discover functional elements in regulatory RNAs or perform strand-specific screens.

7.3 CRISPR Screen Applications: Genetic Mechanisms of Human Disease and Therapeutic Development

7.3.1 CRISPR Screens in Cancer for Synthetic Lethality and Drug Resistance

Over the past few years there has been tremendous excitement surrounding precision medicine approaches for the treatment of diverse cancers [40, 41]. Despite this excitement, there are still many aspects of cancer genetics and therapeutic resistance that are poorly understood. CRISPR screens for cancer functional genomics fall broadly into three major categories: (1) understanding synthetic lethality and identifying potential new therapeutic targets through screening for cancer- and stage-specific dependencies; (2) finding genes that drive resistance or sensitivity to existing targeted therapies; (3) identifying noncoding regulatory elements that influence oncogene expression to provide alternative targeting options in cases where the oncogene itself may not be druggable.

7.3.1.1 Identifying Cancer-Specific Vulnerabilities

Due to different underlying mutational processes and genome instability, cancer cells often evolve different genomic signatures during cancer progression. Characterizing cancer-specific vulnerabilities requires finding mutated proteins or gene expression programs that are essential to proliferation. These identified targets can be candidates for developing targeted therapy.

By applying genome-scale CRISPRn to multiple cancer cell lines, several groups have identified shared essential (core) genes across different cancer types [42, 43]. For each tumor cell line, we can define context-specific fitness genes by subtracting shared essential (core) genes from all essential genes for that tumor. One recent study comparing four cancer types discovered several context-specific fitness genes in glioblastoma, colorectal carcinoma, cervical carcinoma and melanoma [43]. Intriguingly, two different colorectal carcinomas displayed distinct vulnerabilities, highlighting the potential for using a genome sequencing and/or functional genomic screens to stratify patients.
For rare tumors, combining CRISPR screens with patient-derived *in vitro* models can be helpful for correlating functional genomic data with known pathological features and specific genetic mutations (germline or somatic). In a recently established patient-derived cell line for a rare undifferentiated sarcoma, multiple screening approaches (CRISPRn, RNA interference and pharmacologic screens) converged on CDK4 (a cyclin dependent kinase) and XPO1 (a protein involved in nuclear transport) as potential therapeutic targets [44]. One powerful aspect of this study was that the intersection of all three different screen modalities was used to build greater confidence in the genetic hits, suggesting a novel approach to pooled screen validation. In addition to patient-derived *in vitro* models, *in vivo* mouse models have also been employed to understand specific mutations and to characterize multi-cell interactions, such as primary tumor growth and distal organ metastasis. In one type of *in vivo* CRISPRn screen, tumor cells are transduced *ex vivo* with a lentiviral sgRNA library and then the mutant cell pool is transplanted into immunocompromised (or syngenic) mice. Using this approach, a study identified loss-of-function mutations that contribute to primary tumor growth and cancer metastasis *in vivo* by separately analyzing enriched sgRNA targets in different organs [45]. The identified mutations included both well-established tumor suppressor genes, microRNAs (miRNAs) and several novel drivers of metastasis. It was shown that mutations that drive lung metastasis also stimulate primary tumor growth, suggesting that these events are tightly linked for many genetic driver mutations [45]. Another type of *in vivo* CRISPRn screen delivered a sgRNA library using the piggyBac transposase and identified novel tumor suppressor genes associated with liver tumorigenesis [46]. Since it can be challenging with non-virally delivered transposase to limit genomic integration to only a single sgRNA per cell, secondary validation of screen hits is essential to confirm their roles in tumorigenesis.

### 7.3.1.2 Understanding Mechanisms of Drug Resistance

A major obstacle for targeted therapy is drug resistance: When patients are treated with drugs targeting specific oncogenes (such as *BRAF* in melanoma or *EGFR* in non-small cell lung cancer), they often develop resistance to treatment [47]. Genome-wide CRISPRa and CRISPRn screens identified gain-of-function and loss-of-function mutations in BRAF inhibitor-resistant melanoma, and loss-of-function mutations in etoposide-, cytosine arabinoside (Ara-C)- or ATR kinase inhibitor-resistant myeloid leukemias [6, 7, 48, 49]. A genome-wide CRISPRa screen for BRAF inhibitor resistance in melanoma identified potential targets for direct pharmacological inhibition [24]. This highlights a key difference between CRISPRa (gain of function) and CRISPRn (loss of function) approaches. For gain-of-function hits from a CRISPRa screen, it is possible to test established target-specific drugs. In cases where a direct inhibitor is not available, cell lines containing the mutation (or engineered to carry it) can be challenged using a high-throughput drug screen of novel compounds.
For loss-of-function CRISPRn screens, it can be more challenging to translate screen hits into drug targets/strategies. For example, a CRISPRn screen identified CDC25A loss-of-function as driver of resistance to ATR kinase inhibition in acute myeloid leukemia (AML) [49]. A WEE1 (G2 checkpoint kinase) inhibitor could restore the ATR inhibitor’s efficacy in the resistant cells by forcing mitotic entry in CDC25A-deficient cells [49]. Another approach for overcoming drug resistance is to identify multi-gene synthetic-lethal interactions, where resistance stemming from a single loss-of-function mutation is reversed by a second loss-of-function mutation (synthetic lethality). One recent CRISPRn screen evaluated synthetic lethality by delivering two sgRNAs to mutate two genes simultaneously [50]. The study attempted to test 1.4 million possible synthetic-lethal interactions among 73 cancer genes and identified a total of 152 successful pairs demonstrating synthetic lethality. In subsequent combinatorial drug validation studies, the researchers validated roughly 75% of the synthetic lethal combinations discovered. Synergistic cytotoxicity identified in CRISPRn screens can be quite informative and can provide a roadmap for downstream combinatorial drug studies. Similarly, CRISPRa screens can also capitalize on multi-gene targeting to identify resistance genes for combinatorial inhibition.

7.3.1.3 Examining Noncoding Regulators of Cancer Gene Expression

In addition to protein-coding genes themselves, there are many regions of the non-coding genome involved in the regulation of protein-coding gene expression. CRISPRi was used to identify nine distal enhancers within 1 megabase of sequences near MYC and GATA1 oncogenes [17]. MYC is a common oncogenic driver in many different cancers [51] and thus mapping enhancer elements that might increase MYC expression is important for identifying potential therapeutic targets. Additionally, noncoding regulators in T-cell exhaustion was studied with a CRISPRn saturating mutagenesis screen [52]. The study mutated all possible sgRNA sites of nine regulatory sequences near the Pdcd1 gene which codes for programmed cell death protein 1 (PD-1). In the context of cancer immunotherapy, PD-1 inhibition has been approved for a wide variety of different malignancies [53]. By correlating functional regions with putative transcription factor binding motifs, the study suggests possible upstream therapeutic interventions to inhibit immune checkpoint pathways. In general, CRISPR screens can be adapted to detect immune checkpoints or regulatory elements of those checkpoints, providing immunotherapeutic strategies to block T cells from being deactivated by tumor cells. Besides targeting enhancer binding sites, CRISPR screens utilizing saturating mutagenesis or deletion can also detect various other types of oncogenic regulators including long noncoding RNAs (lncRNAs) [30], microRNAs (miRNAs) [54], and other important non-coding regions such as introns and untranslated exons [55].
7.3.2 CRISPR Screens in Infectious Disease

Pathogenic organisms such as bacteria, parasites, and viruses present a major problem for human health around the globe [56]. Pooled CRISPR screens have provided insight into host-pathogen interactions by identifying host factors that facilitate or resist pathogen infections and intrinsic pathogen factors that enhance infection.

Identifying host factors that contribute to pathogenicity is an important step in understanding toxicity and treating bacterial infections. CRISPR screens for host-bacterial interactions tend to focus around two key areas: resistance and sensitizing factors. By treating gene-edited cell pools with bacterial toxins or infectious pathogens, researchers can identify resistance and sensitizing factors through analysis of significantly enriched or depleted genes, respectively. For instance, to study host resistance factors against diphtheria and anthrax toxin, a targeted screen of ~300 genes (including cell surface proteins, and proteins involved in endocytosis, trafficking and cell death) identified four enriched cell-surface receptor genes (PLXNA1, FZD10, PECR and CD81) that confer resistance [57]. Upregulation of genes involved in resistance might protect cells from intoxication. On the other hand, sensitizing factors that facilitate infection can also provide mechanistic insight to pathogenesis. For example, studies have shown that Vibrio parahaemolyticus employs two type III secretion systems (T3SS) to inject its payload [58]. A genome-wide CRISPRn screen in human intestinal epithelial cells used a modified Vibrio pathogen where either T3SS was removed to identify protein modification pathways for pathogen entry that are specific to each T3SS [59]. Down regulation of host factors might provide alternative paths to mitigate cytotoxicity in pathogen infections.

Similarly, to understand specificity of viral-host interaction, multiple CRISPR screens have been used to identify receptors for viral entry and necessary cellular components for viral replication in host cells. Host interactions with flaviviruses and retroviruses are two key examples. Flaviviruses are a family of arboviruses that includes West Nile, Dengue, Zika, and Hepatitis C virus [60–62]. A genome-wide CRISPRn screen revealed seven protective genes in the endoplasmic reticulum associated protein degradation (ERAD) pathway, where loss-of-function confers resistance to West Nile virus-induced cell death but does not block viral replication [63]. To look for shared replication facilitators in host cells, a second genome-wide screen identified and validated signal peptidase complex 1 (SPCS1) as key requirement for flavivirus replication [64]. For viral specific host factors facilitating viral replication, a third genome-wide screen discovered distinct host-dependency factors required for Dengue or hepatitis C virus [65]. Identification of these novel host factors provides new avenues for developing specific antiviral therapies. In addition to flaviviruses, CRISPR screens have also provided insight into retroviruses, such as human immunodeficiency virus (HIV). Although the entry receptors for HIV have been well-characterized (e.g. CCR5 and CXCR4), a genome-wide CRISPRn screen discovered several new dependencies, including tyrosylprotein sulfotransferase 2 (TPST2) and solute carrier family 35 member B2 (SLC35B2) [66]. These two
proteins function in a common pathway to sulfate CCR5 so that it can be recognized by HIV. Loss of either of these proteins and the modifications they impart to CCR5 results in strong protection against HIV, suggesting further targets for controlling viral load.

Relatively few CRISPR screens have been performed in pathogens themselves compared to screens in host organisms. Intrinsic pathogen factors contribute to severity of infections and a classic example is the acquisition of antibiotic resistance. Studies have shown that carbapenem-resistant *Enterobacteriaceae* [67] and methicillin-resistant *Staphylococcus aureus* [68] are resistant to nearly all available antibiotics, suggesting that novel antibiotics or treatment options are urgently needed for combating antibiotic-resistant bacterial infections. CRISPR screens can be implemented to characterize new antibiotics and their mechanisms of action. For example, to test a novel antibiotic MAC-0170636, a CRISPRi screen analyzed all essential genes in *Bacillus subtilis*, and identified undecaprenyl pyrophosphate synthetase (*uppS*), an essential molecule in construction of the bacterial peptidoglycan cell wall, as a key target for the antibiotic [69]. In addition to antibiotic resistance in bacteria, CRISPR screens have been extended to examine intrinsic factors in other types of pathogens, such as parasites. Apicomplexan parasites are one of the leading causes of human parasite infections such as malaria and toxoplasmosis [70]. A recent study used a CRISPRn screen to target all ~8000 protein-coding genes in *Toxoplasma gondii* [70]. The study defined roughly 200 previously uncharacterized fitness genes and identified the claudin-like apicomplexan microneme protein (CLAMP) as an invasion factor in the initiation of infection [70]. CLAMP is essential for parasite infection in fibroblast cells. In malaria, CLAMP knockdown blocks the asexual cycle of the parasite, indicating that insights from the pooled screen could potentially transfer to other pathogens in the Apicomplexan phylum [70].

### 7.3.3 CRISPR Screens for Understanding and Treating Inborn Genetic Disorders

Inborn genetic disorders are diseases caused by inherited or de novo mutations that affect early development. In this area, CRISPR screens have been used to find regulators of hemoglobin switching and novel treatments for mitochondrial disorders.

Hemoglobin disorders, such as beta-thalassemia and sickle-cell anemia, are relatively common. There are >300,000 births each year with severe forms of these diseases, which result from defects in the adult form of hemoglobin (β-globin) [71]. In early development, an alternative, fetal form of hemoglobin is the dominant oxygen carrier. In patients with β-globin defects, it has been shown that natural variants that result in expression of fetal hemoglobin (HbF) prevent severe forms of the disease [72]. Through human genetics association studies, the transcriptional repressor BCL11A was found to block expression of HbF. Using a CRISPRn screen in an intron of BCL11A, an erythroid-specific enhancer region was identified [31].
Mutagenesis of the enhancer phenocopies knock-out of BCL11A and results in re-activation of HbF. For therapeutic gene editing, this erythroid-specific enhancer might be a preferred target since it only reduces BCL11A expression in erythroid lineages. A second study by the same group targeted a noncoding region surrounding \textit{HBS1L-MYB}, which contains single-nucleotide polymorphisms associated with HbF levels and other red blood cell traits. They identified several regulatory elements in this region that control \textit{MYB} expression, which also regulates HbF \cite{33}. Taken together, these screens have identified several different regulatory elements that are essential to the expression of different forms of hemoglobin. For patients with hemoglobin diseases, these studies suggest specific noncoding targets for therapeutic gene editing and also specific regulatory genes that could be inhibited with small-molecule drugs.

Mitochondrial disorders encompass a set of diseases that stem from dysfunctions of the mitochondrial respiratory chain \cite{73}. Over 150 genes have been identified in mitochondrial disease, making it the largest class of inborn errors of metabolism. Despite this genetic diversity, most of the current therapeutic strategies utilize broad vitamin supplementation with limited efficacy \cite{74}. A genome-wide CRISPRn screen used death screening (actively selecting dead cells via Annexin V staining) to identify genes linked to mitochondrial disorders \cite{75}. The study identified 191 genes that already known to play a role in oxidative phosphorylation as well as a handful of previously uncharacterized genes (\textit{NGRN, RPUSD3, RPUSD4, TRUB2, WBSCR16, PYURF, METTL17, TMEM261, N6AMTI}) \cite{75}. Other studies have focused on identifying specific targets in the oxidative phosphorylation pathway to find new therapeutic approaches. A genome-wide CRISPRn screen in a cell line where respiratory chain function was impaired (either by antimycin or pyruvate removal) identified the Von Hippel-Lindau (VHL) factor as a potential suppressor of mitochondrial disease. VHL was previously described as a key regulator of cellular hypoxic response, linking the hypoxia pathway with mitochondrial metabolism \cite{76}. The protective effects of VHL knock-out was further validated \textit{in vivo} in zebrafish. In a mouse model of Leigh syndrome, hypoxia treatment ameliorated a respiratory chain defect in which complex I is disrupted and extended lifespan by over threefold \cite{74}. A separate study combined a chemical screen with a genome-wide CRISPRn screen to identify factors that could rescue defects in complex I of the mitochondrial respiratory chain. The chemical screen identified I-BET 525762 as a bromodomain protein inhibitor, and the CRISPRn screen revealed that the target of the inhibitor was the bromodomain containing protein 4 (BRD4) \cite{77}. Ablating BRD4 increases oxidative phosphorylation and, here, the complementary drug screen provided additional support for this hit. Both screens suggest that inhibiting the activity of BRD4 might help the mitochondria compensate for defects in complex I. Overall these studies highlight the potential for new therapeutic approaches and demonstrate that mitochondrial disorders require treatments to be tailored for specific genetic lesions or specific impairments to respiratory chain complexes.
7.4 Conclusion and Future Perspectives

In order to develop new therapies for complex diseases, a key challenge is to identify genes and other functional elements in the genome involved in pathogenesis. With new targeted gene editing technologies, large-scale, pooled genetic screens in human cells are significantly easier than with alternative approaches (e.g. transposons, retroviral insertion, chemical mutagenesis). To date, most pooled screens have focused on probing one target per cell but future screens can take advantage of multiplexing to probe multiple genome targets in a combinatorial fashion. Combinatorial approaches can be useful in cancer and infectious disease in the context of synthetic lethality to identify optimal multi-drug cocktails, and also in inborn genetic disorders to identify background-specific modifiers for disease severity and therapeutic efficiency. With respect to precision medicine, future CRISPR screens could be performed in patient-derived cell lines to identify targets specific to the patient genetic background or to perturb specific gene variants.

In addition to gene targets, there is tremendous interest in understanding how non-coding regulatory regions influence gene expression, given that most common-disease-associated variants are in noncoding regions [78]. A key problem going forward for high-throughput pooled screens is to find screenable (cell autonomous) phenotypes for complex diseases. Traditionally, pooled screens have employed survival phenotypes (e.g. resistance to a drug or a pathogen) but many disease-relevant phenotypes are subtle or difficult to analyze in a pooled format. Despite these challenges, new advances in CRISPR pooled screening, such as recent work to combine pooled editing with single-cell readouts of RNA, DNA or genome state [79–83], deletions to perturb larger regions of the genome [30, 84], and new effector domains for manipulating epigenetic states [11, 27], will improve our understanding of the genetic basis of disease and help identify new therapeutic targets for treating these diseases.

References

Material and Methods

1. Material and methods...


