Target discovery and drug development
what comes after CRISPR?

For an area of research as important and impactful as drug discovery, it is surprising how often efforts in this field are directed by short-term trends. We have seen an abundance of new tools and platforms over recent years, but how should they be enabled to plot a path from target ID screen to new medicines available in the clinic?

By Benedict CS Cross

Over the last five years, pharma and biotech companies alike have spent a huge amount of energy and investment pivoting their pipelines to align with the seismic influence of CRISPR-based technologies. This dramatic shift is a testament to the tremendous influence of this transformative technology, from discovery through to treatment. Early results are already starting to validate the wisdom of the CRISPR-rush and the pace of adaptation and development is likely to keep the fires burning for quite some time to come.

Nevertheless, as we have learnt more about CRISPR and related gene editing technologies, there are already some overlapping and hard lessons from RNAi – success is possible, but the path from target discovery to the clinic is by no means straightforward. In this context, it is prudent to start now to look beyond CRISPR – what is the most effective use of this technology, and how can we augment it where it is weak?

Screening tools that have been deployed for loss-of-function can already be adapted and repurposed to alternative perturbations, but even these tools cannot interrogate the subtle biological processes that can govern clinical response. Fortunately, new screening tools exist that can find both the targets and the mechanisms that modulate complex molecular interactions. Here, I will discuss the changing drug discovery landscape and look towards a bright future of multi-modal screening and drug development.

Target ID – lessons from RNAi

Drug candidate attrition and late-phase failure remain a real and sustained challenge for the pharmaceutical discovery efforts of the research community. To compound issues, the discovery of novel targets through traditional means has yielded a striking dependence on a tiny proportion of the genome for the development of new drug development pipelines[1]. Only around 700 human biomolecules are as yet successfully targeted for therapeutic intervention, representing less than 1% of the proteome[2], and still dominated by a core set of families in kinases, ion channels and GPCRs. There is a self-evident requirement to identify the most robust and rapid ways to determine suitable new target space and mechanisms of drugability across all human diseases. Ideally, such campaigns should be high-throughput and unbiased, sampling as much biological space as possible to maximise the diversity of pathways and biochemical architecture under interrogation.

Since the dawn of RNAi, functional genomics has become a central pivot for massive discovery

References

1 Terry, C and Lesser, N. Unlocking R&D productivity: Measuring the return from pharmaceutical innovation 2018. Deloitte Cent. Heal. Solut. (2018).

2 Santos, R et al. A comprehensive map of molecular drug targets. Nat. Rev. Drug Discov. 16, 19-34 (2016).

3 Mohr, SE, Smith, JA, Shamu, CE, Neumüller, RA and Perrimon, N. RNAi screening comes of age: Improved techniques and complementary approaches. Nature Reviews Molecular Cell Biology (2014) doi:10.1038/nrm3860.

4 Sims, D et al. High-throughput RNA interference screening using pooled shRNA libraries and next generation sequencing. Genome Biol. (2011) doi:10.1186/gb-2011-12-10-r104.

5 Bassik, MC et al. Rapid creation and quantitative monitoring of high coverage shRNA libraries. Nat. Methods (2009) doi:10.1038/nmeth.1330.

6 Schlabach, MR et al. Cancer proliferation gene discovery through functional genomics. Science (80-.). (2008) doi:10.1126/science.1149200.

7 Tsherniak, A et al. Defining a Cancer Dependency Map. Cell (2017) doi:10.1016/j.cell.2017.06.010.

8 McDonald, ER et al. Project DRIVE: A Compendium of Cancer Dependencies and Synthetic Lethal Relationships Uncovered by Large-Scale, Deep RNAi Screening. Cell (2017) doi:10.1016/j.cell.2017.07.005.

9 Bhinder, B and Djaballah, H. A decade of RNAi screening: Too much hay and very few needles. Drug Discov. World (2013).

10 Kampmann, M, Bassik, MC and Weissman, JS. Integrated platform for genome-wide screening and construction of high-density genetic interaction maps in mammalian cells. Proc. Natl. Acad. Sci. (2013) doi:10.1073/pnas.1307002110.

Continued on page 12

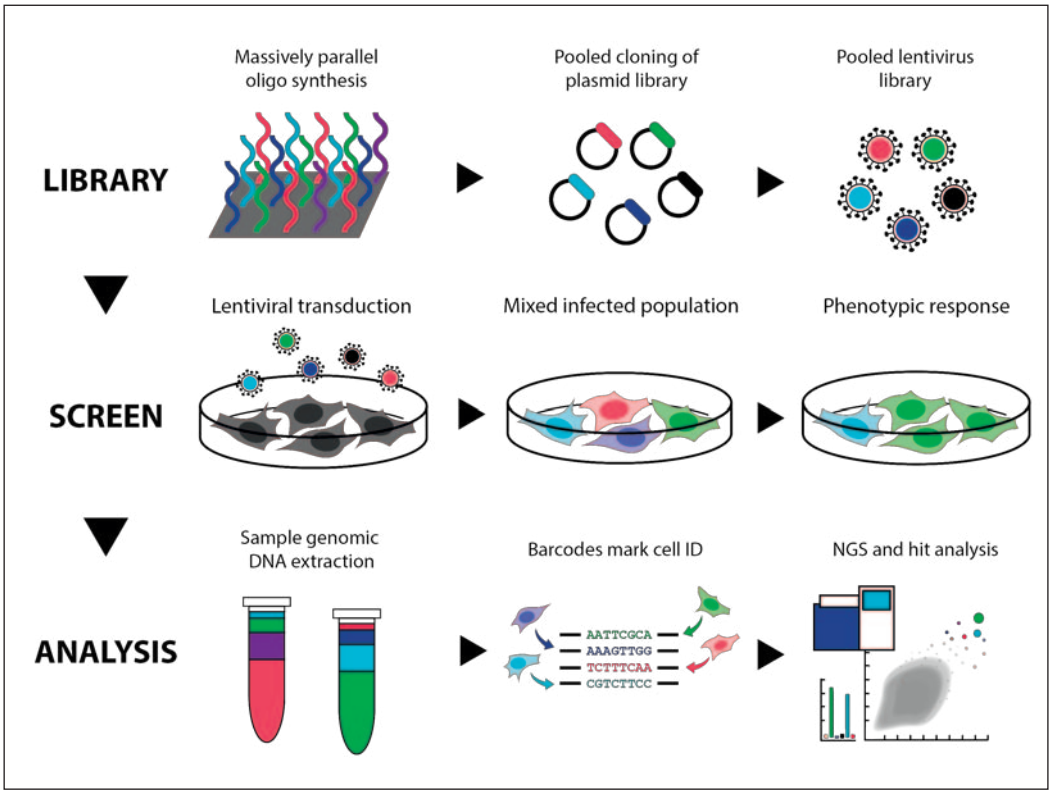

Figure 1: Pooled screening technologies and the practical steps for target ID with these approaches

efforts in both basic biology and drug development. This technology equilibrated mammalian research with the simple model organisms most commonly used for high-throughput genetic approaches, and human studies no longer had to rely on complex or intractable association studies. RNAi exploits innate anti-viral cellular pathways to selectively silence individual transcripts, and can be programmed by scientists to target specific mRNAs with systematic sequence-directed degradation. This relatively simple tool was initially enabled by rapid and straightforward synthetic chemistry, but the automation required for large-scale studies lagged behind and data quality from these studies varied[3]. Drug discovery teams needed a new way to use this tool, and to be able to explore genome-wide analysis with a much higher quality level. Deep sequencing innovations provided the route forward[4]. The drive to bring costs down in next generation sequencing meant that access to massively parallel DNA sequencing could be achieved for research in completely new areas. Researchers used pooled oligo synthesis to generate genome-wide libraries of RNAi targeting sequences and then extracted quantitative, phenotypic data from their experiments using next-generation sequencing approaches[5,6] (**Figure 1**).

Now biologists were free to explore cell systems with genome-wide loss-of-function tools. This spectacular advance was pursued with feverish enthusiasm, and vast resources were poured into RNAi-based pooled screening in both academia and industry alike. Drug discovery groups in biopharma started to base their target discovery efforts on this tool to try to fill their pipelines. A large-scale, pan-cancer cell panel screening campaign was initiated by multiple groups[7,8]. Committees were established to harmonise data analysis approaches, and carefully optimised targeting sequences and custom algorithms were developed to predict on- and off-target effects.

Over the course of the 10 years since, hundreds of screens have been conducted in a huge diversity of disease systems, but perhaps the only clear conclusion from all of this effort was that neither the penetrance nor the precision of the tool was sufficient to generate clean data for robust target ID 9. The drug discovery industry could not rely on the data from these screens, and enthusiasm waned dramatically. Through necessity, the field moved to ultra-complexity – driving the number of replicates and shRNA hairpin number to the maximum the technology could cope with. This provided more promising results[10,11], but it was clear that

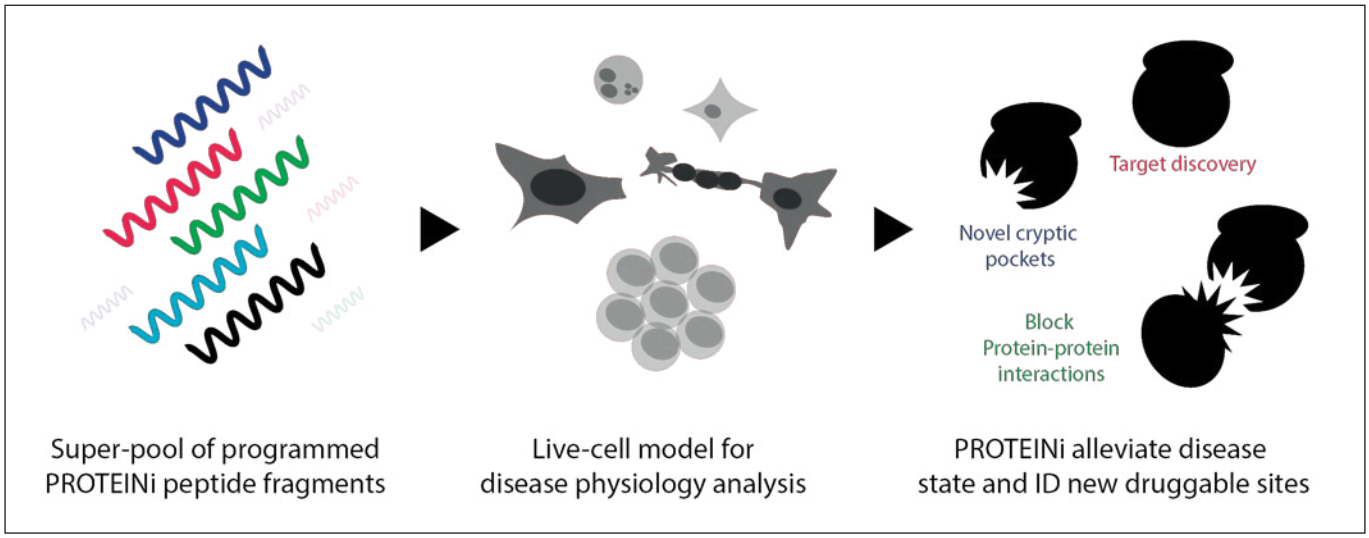

Figure 2
PROTEINi uses a massively diverse collection of 3D shapes, programmed by DNA-encodable peptide fragments. These tools are used in a live cell, physiological disease model to identify novel targets and mechanisms to treat disease

improved technologies would be needed to realise the full potential of pooled screening approaches.

Pooled screening with CRISPR – a step in the right direction

The discovery of the CRISPR-associated enzymes, an RNA-guided gene-editing tool, has been nothing short of revolutionary for biologists, and may yet yield a similar impact as a therapeutic in its own right[12]. In the case of target ID, it has represented a very clear increment in quality from the extant RNAi tools. CRISPR uses a programmable nucleic acid sequence to direct a large bi-nuclease to a specific locus. CRISPR now comes in a wide variety of adapted versions[13], but in its most commonly-used form it is able to ablate gene expression completely from the targeted genomic locus[14].

The most immediate opportunities for drug discovery researchers and for target ID required the adaptation of the technology to genome-wide screening. This owes a substantial debt to RNAi, where the tools were rapidly and directly translated to generate a pooled CRISPR-based screening approach. In principle, the suitability of a CRISPR mechanism of perturbation to pooled screening is much greater than for RNAi, owing to the likelihood for clustering of effect sizes and response. This means that each time you target a gene with an sgRNA, the same phenotypic response should be elicited, rather than for shRNA where silencing efficiencies represent a broad distribution of effects[13]. This has technical advantages for the detection and statistical significance of the measurement, and coupling this feature to high effect rate through genetic ablation provides an excellent window for observing meaningful phenotypic effects.

With technologies as dramatically impactful as CRISPR, it is easy to get lost in the vast potential and lose sight of the pragmatic requirement to carefully evaluate the appropriate application. There are undoubtedly some truly exciting outcomes to arise from use of this tool in understanding biological mechanisms, and in medicine, but is it a panacea for target ID?

The system is not perfect and even in parallel studies of response from two ostensibly identical screening analyses, hit overlap is restricted to a few top genes and relative ranking is often incomplete between studies[15,16]. This can likely be attributed to modest differences in the approaches used and the libraries of guide RNAs in the study, and of greater importance is an evaluation of the tool's ability to discover novel targets. The emerging evidence is of a higher than expected false negative rate. In a systematic analysis of recent large-scale cell panel screening with CRISPR knock-out, it was estimated that up to half of all constitutively expressed genes have never yielded a phenotype in any screening campaign[17]. This phenomenon is likely explained by the high level of complexity and redundancy in the human genome, resulting in functional buffering which masks the effect of individual genetic perturbations. For drug discoverers, this is a vital lesson – when using a single screening tool to populate novel target pipelines, careful experimental validation is necessary when interpreting hits arising from campaigns.

Gene knock-out yields a theoretical maximum effect size for the loss of target function. This feature is both the technical strength and simultaneous weakness for target ID in drug discovery.

How many drugs will be able to recapitulate this hypomorphic effect size? It is rare that 100% loss-of-function can be achieved by a small molecule inhibitor, and knock-out screening may result in a pool of unactionable targets for drug development. One promising way around this is to conduct parallel silencing and gene editing-based screening to skew the hit nomination in favour of those targets which are robustly discovered even with modest penetrance of perturbation[18]. These studies and others[19,20] offer the first glimpses of multi-modal screening and present the beginnings of a new strategy for target ID. Perturbation-based approaches are necessarily tied to the relatively simplistic, binary interrogation of function and do not directly allow the analysis of the more complex biological interactions. Thus, while loss-of-function tools provide powerful mechanisms to identify biological response pathways, they are not always well suited to discovering novel druggable modules.

Future target discovery now

New screening mechanisms are being developed all the time, and many of these are positioned to substantially augment the current commonly-used tools. One very promising approach is to adapt the existing high-throughput screening infrastructure to more directly explore the drug-like space. DNA-encoded libraries of small molecules using a sequencing-based readout and phage-display can provide analysis of many millions of fragments, but neither of these are currently compatible with live-cell mammalian screening systems, which is crucial for concise conversion of discoveries into drugs.

New screening approaches

The tools used for shRNA and CRISPR-based pooled screening have been adapted by PhoreMost to study peptidyl drug-mimetics in live cells and at vast scale (**Figure 2**). This innovative approach overcomes some of the challenges faced by perturbation-based screening, and by maximising the capacity of the pooled screening apparatus, a massive exploration of potential drug space can be accessed.

The fully programmable PROTEINi® (Protein interference) approach opens up the opportunity to explore target response and discovery by directly interrogating target engagement in an unbiased screen. This is a crucial differentiator to straightforward perturbation analysis, as it allows not only the examination of loss-of-function, but also the ability to more intricately survey the nuances of

Continued from page 8

11 Morgens, DW, Deans, RM, Li, A and Bassik, MC. Systematic comparison of CRISPR/Cas9 and RNAi screens for essential genes. Nat. Biotechnol. 34, 634-636 (2016).
12 Stadtmauer, EA et al. CRISPR-engineered T cells in patients with refractory cancer. Science (80-.). (2020) doi:10.1126/science.aba7365.
13 Le Sage, C, Lawo, S and Cross, BCS. CRISPR: A Screener's Guide. SLAS Discovery (2019) doi:10.1177/2472555219883621.
14 Wang, H, La Russa, M and Qi, LS. CRISPR/Cas9 in Genome Editing and Beyond. Annu. Rev. Biochem. 85, 227-264 (2016).
15 Le Sage, C et al. Dual direction CRISPR transcriptional regulation screening uncovers gene networks driving drug resistance. Sci. Rep. 7, (2017).
16 Shalem, O et al. Genome-scale CRISPR-Cas9 knockout screening in human cells. Science (80-.). 343, 84-87 (2014).
17 Dede, M, Kim, E and Hart, T. Biases and Blind-Spots in Genome-Wide CRISPR Knockout Screens. bioRxiv (2020) doi:10.1101/2020.01.16.909606.
18 Deans, RM et al. Parallel shRNA and CRISPR-Cas9 screens enable antiviral drug target identification. Nat. Chem. Biol. 12, 361-366 (2016).
19 Jost, M et al. Combined CRISPRi/a-Based Chemical Genetic Screens Reveal that Rigosertib is a Microtubule-Destabilizing Agent. Mol. Cell 68, 210-223.e6 (2017).
20 Le Sage, C et al. Dual direction CRISPR transcriptional regulation screening uncovers gene networks driving drug resistance. Sci. Rep. 7, (2017).
21 Carvalho-silva, D et al. Open Targets Platform: new developments and updates two years on. 47, 1056-1065 (2019).

Continued on page 13

biological activity that modulate regulatory pathways. This is a space not captured by other technologies used for drug target discovery and is a major mechanism for developing new therapeutics that can tackle otherwise undruggable diseases.

PROTEINi screening is therefore positioned to substantially amplify the existing target ID space, and it might also be able to rescue orphaned hits from orthologous screening tools. To this end, an important feature of perturbation-based screening campaigns is the discovery of highly significant hits which fall into a traditionally undruggable classification. Often these hits are simply ignored by researchers as they do not have obvious routes to drug development, in favour of hits in known pathways, but this response defeats the purpose of conducting the screen in the first place. Using programmable, high shape-diversity focused screens against these targets with tools such as PROTEINi could unlock functional mechanistic sites or cryptic druggable pockets in those targets, providing a new and direct route to drug development (**Figure 3**).

Data mining and AI

Additional tools for new target discovery include new collaborative assimilation models and these hold great promise for new target ID. The Open Targets initiative is an example of a new wave of platforms from which valuable data can be mined[21]. The contributors to the tool, which include academics as well as industry teams, pool their data and a custom set of algorithms and analyses is conducted to build target-disease associations from publicly-available information. As with any omics-like tool, it is subject to the quality of the information on the input side, but does represent a more co-operative future for drug target ID.

Open resource tools can also prove useful for machine learning and AI outcomes, and there has been an explosion of companies founded on this approach in the last five years. It is yet to be determined what the AI gold rush will ultimately contribute, but the first reports of AI-designed drugs are beginning to emerge. For the purposes of target discovery, AI is currently acutely dependent on largely public *in vitro* datasets for training, and similar *in vivo* response and toxicity data is simply not available at the scale required. Thus, it is not clear yet how easily AI discovery mechanisms will translate their outputs into clinical efficacy. Its contribution to our understanding of protein interactions and dynamics might be more tractable, as the *in vitro* datasets needed to feed the computation are already in existence at the scale and detail required.

Cell models and phenotypic measurements

In vitro screening tools for target ID require robust model systems within which to operate, and screeners have to determine whether a simple cell line-based approach is adequate or if a more complex physiological model is required. In the case of cancer target discovery, a wealth of cell panels are available, and while appealing due to their accessibility, in many cases these systems provide more noise than desired. Conflating mutations, complex genomic and epigenomic landscapes within each tissue and cell model can cause substantial barriers to hit quality. Isogenic systems offer a cleaner background for screening, but recapitulating discoveries from these backgrounds into patient models can be challenging. Screening, therefore, often needs to be conducted in multiple cell backgrounds simultaneously to provide sufficiently translatable hit discovery.

Outside of oncology there is a major requirement for screen-ready physiological cell models, and these are now starting to become available at the scale needed for unbiased target ID. The innate malleability of induced pluripotent stem cell (iPSC) systems, and improvements in their biomanufacturer, is a valuable new development. However, it is possible that realisation of the new discovery potential of these cell systems will require more complex phenotypic analysis of perturbation response. Here, single cell analysis can provide a crucial additional degree of resolution, where coupling ultra-rich data to target ID screening is already possible in some forms, albeit at a smaller scale than ideally required[22-24].

Outlook: multimodal screening to drug the undruggable

With all of these new platforms and tools available, how does one plot a path from target ID screen to new medicines available in the clinic? It is increasingly clear that each tool, from AI to PROTEINi through CRISPR, has selective advantages and predispositions to discovery in different spaces. Thus, the most promising routes to drug target development are likely to be through a combined and pragmatic application of each in accordance with their respective merits.

In this pluralistic model of multimodal screening, it is particularly important not to overlook one of the hallmarks of target progression: druggability. This feature of new targets has historically been a major criterion with which to judge and prioritise key hits arising from target ID campaigns, resulting in the loss of potentially invaluable, but seemingly intractable, targets. Drug-mimetic

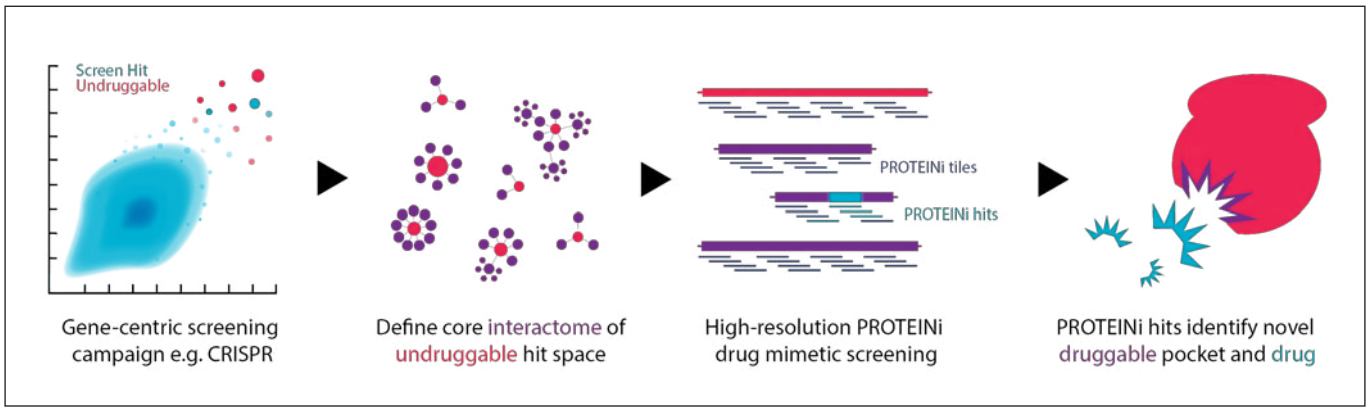

Figure 3: Maximising the value of perturbation-derived target discovery campaigns, such as with CRISPR or GWAS data. Many screens yield targets which cannot be drugged by traditional approaches. New technologies such as DNA-encoded libraries, AI and PROTEINi could be used to rescue the best hits by designing or deriving new drugs against the hits

screening, such as with PROTEINi, provides a potentially direct route to capture targets which are necessarily pre-selected for ligandability, decreasing the time spent on otherwise orphaned targets and expanding the target space available.

Other tools are also providing new mechanisms for drug development which can influence the traditional target ID rules. Adaptation of the ubiquitin proteasome system for target degradation allows drug developers to unlock targets that might otherwise have been considered undruggable, while therapeutic gene editing approaches are also starting to show promise. These developments are important since they lower the barriers in the traditional criteria for drug target progression and allow the target ID activity to be conducted in an unconstrained manner, bringing better targets to market faster. The route for pharmaceutical and biotech companies to populating their target pipelines is inextricably linked to their ability to conduct innovative target ID screening. What comes after CRISPR? A multitude of tools used in parallel, to properly capture the total target space. DDW

***Benedict CS Cross** is Senior Director of Technology at PhoreMost Ltd. He is a geneticist and biotechnologist, with a background in both reverse chemical genetic and forward genomic screening. In 2014 he founded the CRISPR-centric functional genomic screening department at Horizon Discovery Ltd before moving to PhoreMost in 2019. PhoreMost has developed SITESEEKER®, a novel phenotypic screening technology, to uncover cryptic druggable pockets in novel disease-modifying targets.*

Continued from page 12

22 Datlinger, P et al. Pooled CRISPR screening with single-cell transcriptome readout. Nat. Methods 14, 297-301 (2017).
23 Dixit, A et al. Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. Cell 167, 1853-1866.e17 (2016).
24 Adamson, B et al. A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the Unfolded Protein Response. Cell 167, 1867-1882.e21 (2016).