Checklist for CRISPR Screen Design [template]

Jesse Engreitz – June 2021

Biological question:

Which distal elements regulate *GATA1* expression?

Goals of pilot experiment:

Establish CRISPRi system works well in K562 cells
Demonstrate feasibility of mapping quantitative effects on GATA1 expression
Validate the specificity of the GATA1 FlowFISH assay.

Cellular model and phenotypic readout:

FlowFISH for GATA1 in K562 cells

Cellular model. Technical questions to answer about your cell line or cellular model:

- Which cellular model? Is it appropriate for the biological question?
  K562 cells — erythroid cell line where GATA1 is essential

- How will I deliver CRISPR perturbations? Can I establish a stably-expressing CRISPR line?
  For CRISPR: lentiviral dox-inducible KRAB-dCas9 stably integrated into K562 cells
  For guides: sgOpti with puro selection

- How quickly does this cell line silence lentiviral constructs?
  Yes, over weeks; need to periodically re-enrich for BFP+ cells

- How many cells can I easily infect with gRNA lentivirus?
  Easy to do millions

- How many cells can I possibly infect with gRNA lentivirus?
  Possible to infect 100s of millions

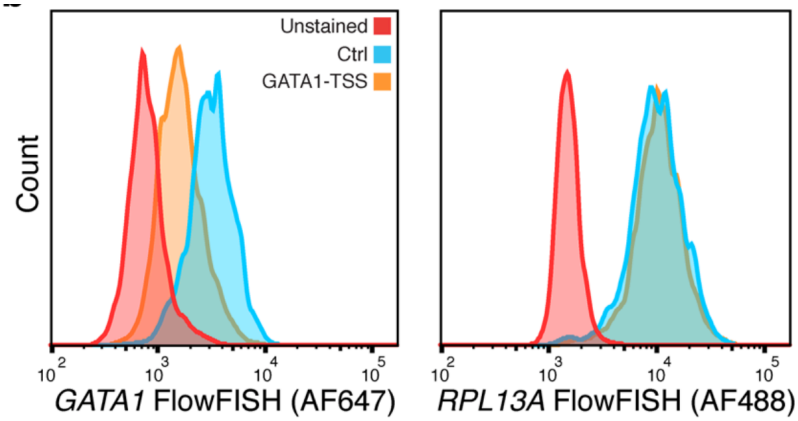
- Is gene knockdown as efficient as expected in this line? (*e.g.*, >90% knockdown with *ENO1* gRNA)?
  Confirmed

Phenotypic readout.

- What is the phenotype or readout used to measure the effects of gRNAs?
  FlowFISH for GATA1

- What positive controls will I use to test the assay?
  Knockdown of GATA1 by gRNAs targeting the promoter, or siRNAs against GATA1

- What is the signal-to-noise ratio for positive vs negative controls (*e.g.*, gRNAs, or +/- stimulation)? How quantitative is the assay?



Quantitative, based on individual FlowFISH experiments using gRNAs targeting GATA1 promoter, enhancer, and negative control gRNAs

- How reproducible is the assay?

Verified by doing technical replicate FlowFISH experiments + flow readout

- How efficient / accurate is my guide readout + assignment (*e.g.*, for Perturb-seq or other single-cell screens?)

Guide readout is by PCR of gRNA amplicons, which is fairly efficient

Scale. The possible scale of a pooled CRISPR screen depends on a number of technical considerations.

- What is the maximum size of gRNA pool that I can ***easily*** profile for a pilot study?
- What is the maximum size of gRNA pool that I can ***possibly*** profile for a full-scale study?

Scale calculation:

The current FlowFISH protocol involves processing ~10-20M cells per tube, and results in 5-10M cells remaining after all of the steps. It is possible to process up to 24 tubes per round. So, an easy scale of experiment would have ~5 M cells to work with. The largest size experiments at the moment would have ~200M cells.

Assuming that we want to aim for >1,000x coverage (*i.e.,* 1,000 single cells per gRNA — this seems sufficient for a typical FlowFISH screen), an easy experiment would have 5M cells / 1000 = 5,000 gRNAs, and a large experiment would have 200M cells / 1000 = 200,000 gRNAs

This scale is feasible for K562 cells — obtaining 200M infected cells is possible (would require infecting ~2 billion cells assuming 10% MOI in multiple large spinner flasks), and obtaining 5M infected cells is fairly easy.

For CRISPRi tiling experiments, we now know we want to include 15 gRNAs per element (see rationale below). For example, a small experiment of 5,000 gRNAs would be able to study ~300 elements + 500 negative control gRNAs.

Designing the gRNA Pool:

Question:	Analysis strategy:	gRNAs to include:
What is the distribution of null effects for non-targeting gRNAs in the assay? How tightly or widely distributed are they? (affects power of the assay) Are there frequent off-target effects of CRISPR gRNAs on this phenotype?	Assess the distribution of effect sizes for many negative control gRNAs that do not target any region in the human genome	Random scrambled **non-targeting gRNAs** — which are designed such that they should not target any site in the reference genome. In many of our screens we have included a selection of random gRNAs that the Weissman Lab created in one of their initial CRISPRi genome-wide screening paper [Gilbert et al., Cell 159, 647–661 (2014).]. These gRNAs had already been filtered for low off-targets (at least in human genome? Check the paper and edit this). File is now part of the CRISPRiTilingDesign github repository (data/Weissman*)
Does gRNA delivery anywhere in the genome affect the phenotype, *e.g.* due to effects of cutting and DNA repair on cell state?	Conduct a T-test comparing gRNAs targeting random regions of the genome to non-targeting gRNAs	gRNAs targeting random regions of the genome (**safe-targeting gRNAs)** As of 2020, Engreitz Lab has been using the ENCODE set of “safe-targeting” loci originally selected by Josh Tycko. These are included in the github repo
Are there any unexpected regional effects of the particular CRISPR effector used in this assay? For example, for cutting screens, does the process of cutting and DNA repair (rather than the resulting indel) have an effect on a nearby gene? In rare events, cutting within a few megabases of the end of a chromosome sometimes leads to loss of the end of the chromosome. Is this frequent enough to affect your readout? For CRISPRi screens, does CRISPRi delivery anywhere in a given locus affect a given gene?	Conduct a T-test comparing gRNAs targeting nearby regions of the genome to non-targeting gRNAs	gRNAs targeting regions of the genome that are close to your regions of interest (*e.g.*, a non-accessible intergenic region 5-100Kb away from the gene or enhancer of interest)

Are there any other ways to generate negative control sequences that more perfectly match my targeted set of sequences?	Conduct a T-test comparing gRNAs targeting these negative control sequences vs non-targeting gRNAs	Examples: For a CRISPRi screen targeting DNase peaks in K562 cells, we could select as controls sequences that are DNase peaks in other cell types — which should be more closely matched in terms of overall sequence content and proximity to genes. For a CRISPRi screen targeting promoters: - Promoters of genes that are not expressed - Promoters of genes that are not relevant to your phenotype
What is the range of effect sizes for real elements on the desired phenotype that we should see in the assay?	Perturb a range of positive control elements and quantify the effect sizes versus non-targeting gRNAs	Examples of positive control gRNAs: For CRISPRi-FlowFISH: gRNAs that directly target the promoter of the gene for FlowFISH For CRISPRi phenotypic screens: gRNAs that target the promoters of genes already known to control that phenotype in this cell line (*e.g.*, from previous publications) For Perturb-seq studies: gRNAs that target genes that should have dramatic effects on transcription (*e.g.*, perturbing TP53 will affect expression of many cell cycle genes)
Is the CRISPR perturbation method working well?	In addition to the row above, additional types of gRNAs can be used as internal positive controls, especially in the case where there are not good positive control genes for the phenotype of interest	Examples: For a CRISPRi-FlowFISH screen, gRNAs targeting known essential genes (*i.e.*, that assess cell proliferation) could be used and assessed for depletion before and after activation of the CRISPR machinery. gRNAs that have been previously used in other screens (and are known to be effective at knocking down the target gene, for example) are very useful positive controls

Pilot screen gRNA Pool contents:

Category	**# gRNAs**
300 DNase peaks near GATA1 x 15 gRNAs per peak	4500
Positive control: gRNAs targeting GATA1 promoter	15
Safe-targeting negative control gRNAs	250
Non-targeting negative control gRNAs	250
Total	5015