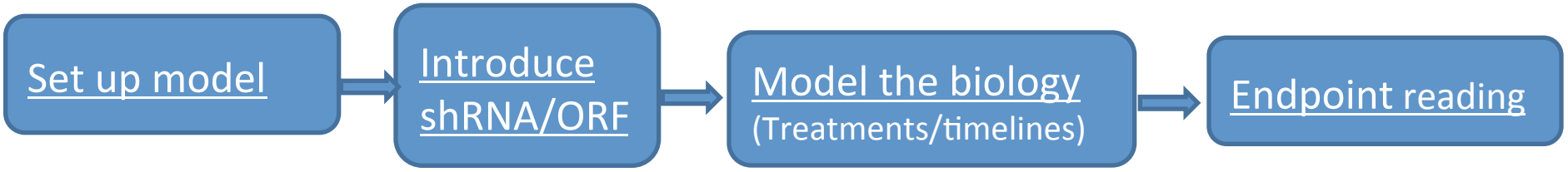
Optimizing and executing shRNA and ORF screens

Arrayed screens & pooled screens for RNAi and ORF screens

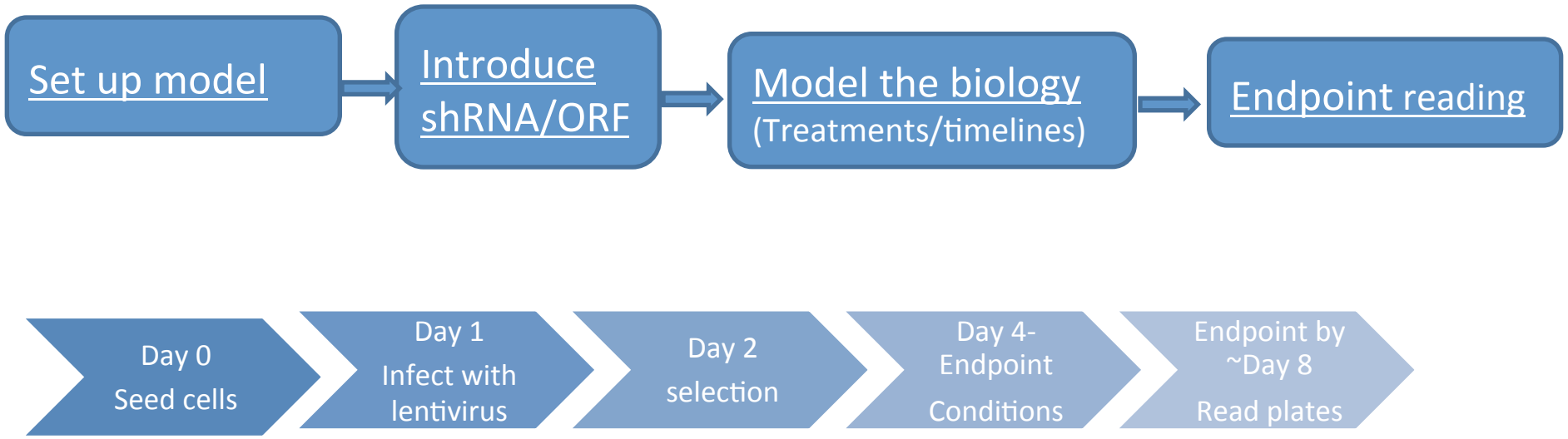
Arrayed:
Multiwell plates, one shRNA per well

Pooled:
Each cell in mixture gets different shRNA or ORF

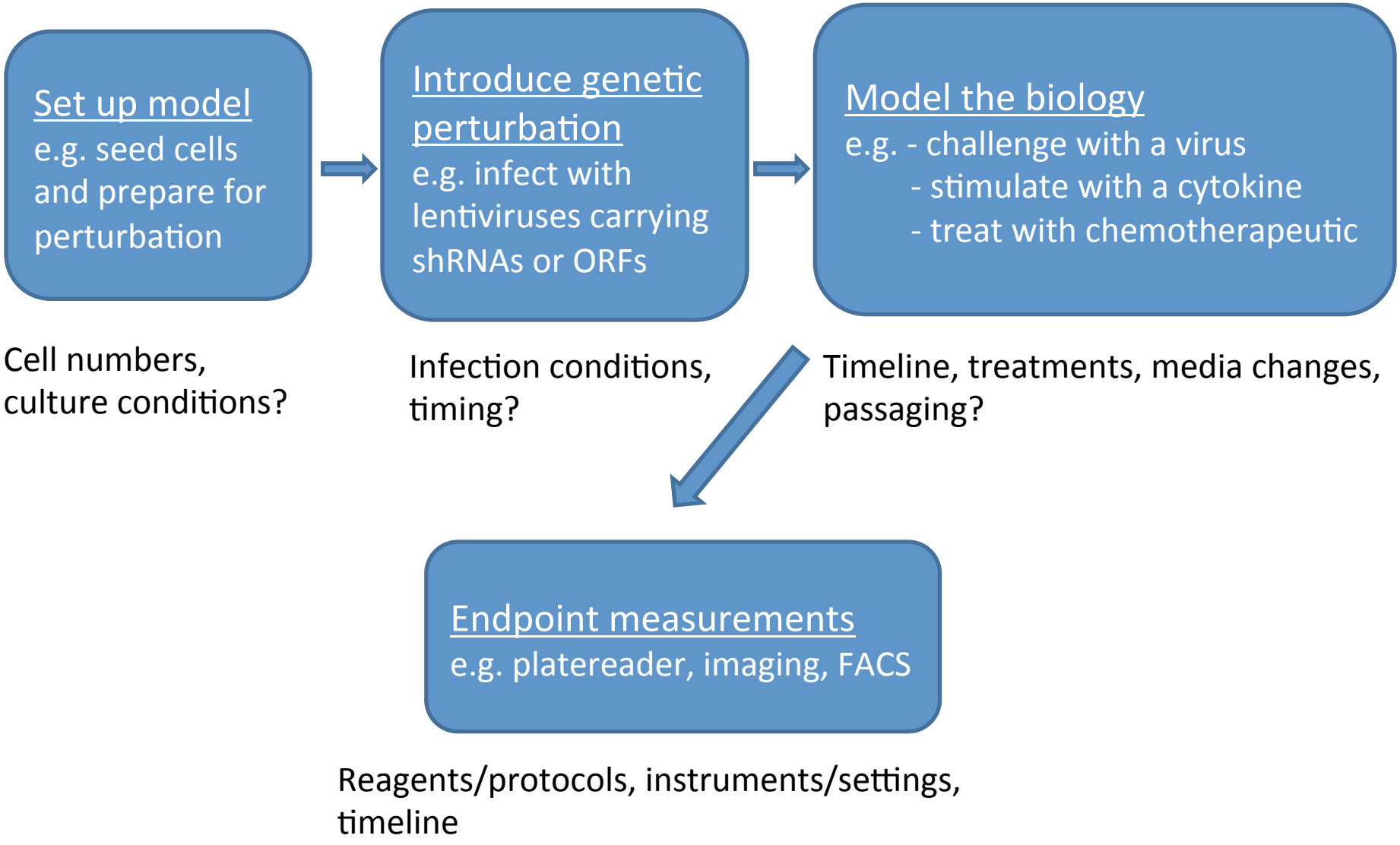
Some issues are similar, some different: we'll start with arrayed screens



Typical timeline for arrayed lentiviral shRNA/ORF screens



Assay Development: Many variables - many may matter!

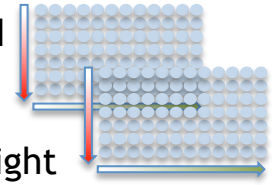


How to optimize many interdependent variables?

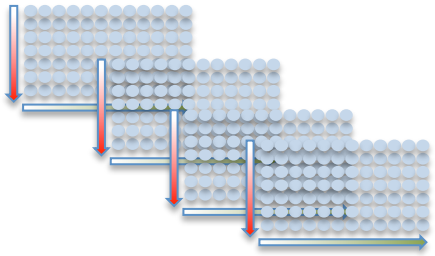
Ex. Just for a subset of infection condition variables below…..

Virus removed after 4 hours

Virus on overnight

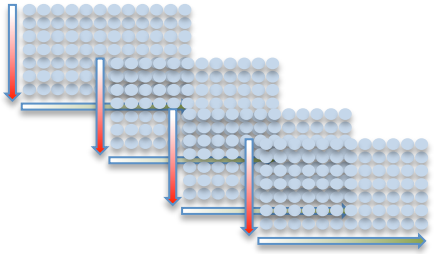


X



Range of cell seeding densities

X

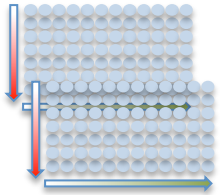


Range of virus doses

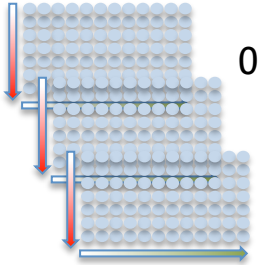
X

No spin

30 minute spin



X



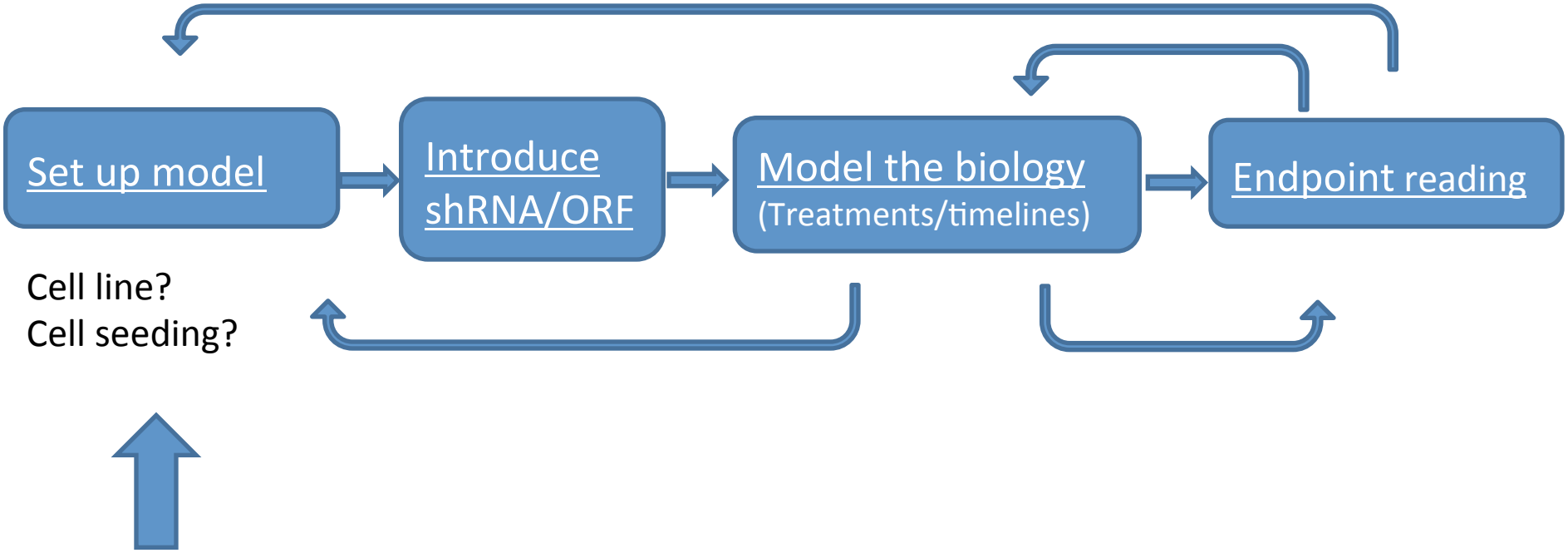
0 ug/mL polybrene

4 ug/mL polybrene

8 ug/mL polybrene

HOW TO DO with finite time and effort?

Typical Series of Experiments to optimize a screen

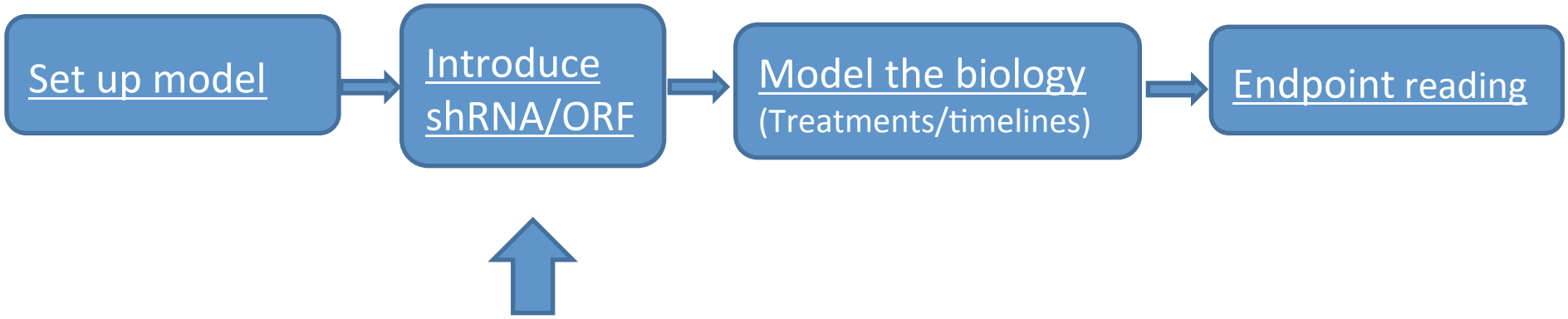


1. Pick cells for screen development – based on biology and logistics
   One or more for primary screen, others for follow-up

2. Seed them at 2-4 reasonable (guessed) densities
   (growth rate, time until endpoint…)

Use this density to do initial optimization of infection, selection.

Optimize library infection conditions

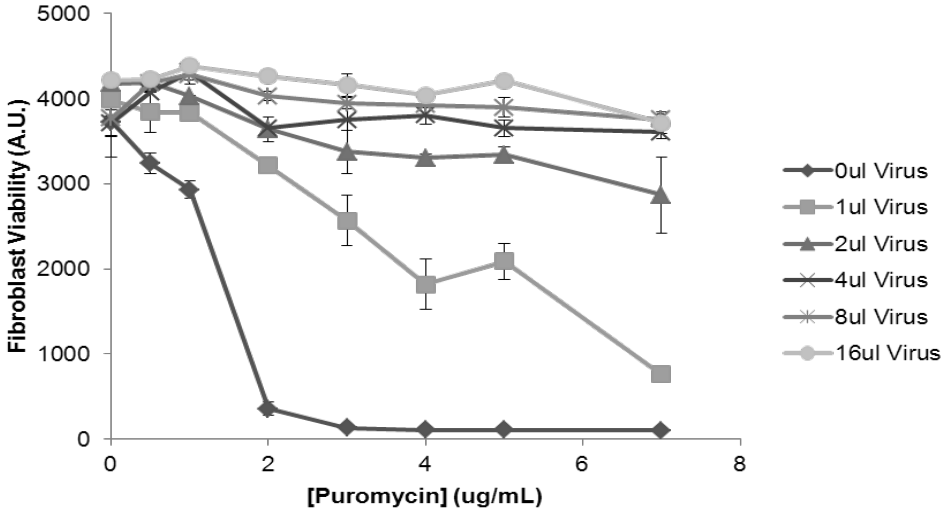


Selection of infected cells

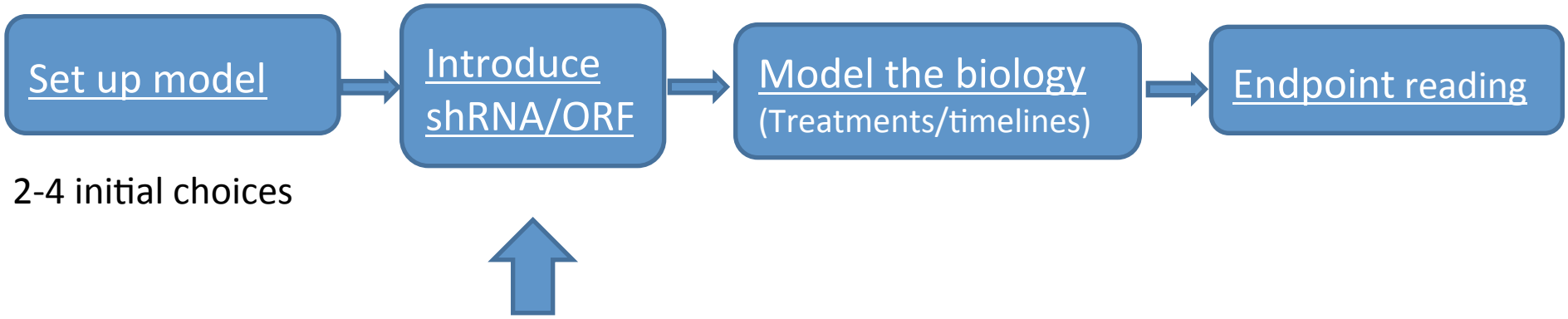
needed for subsequent experiments:

What's minimum puromycin that will kill un-protected, un-transduced cells?

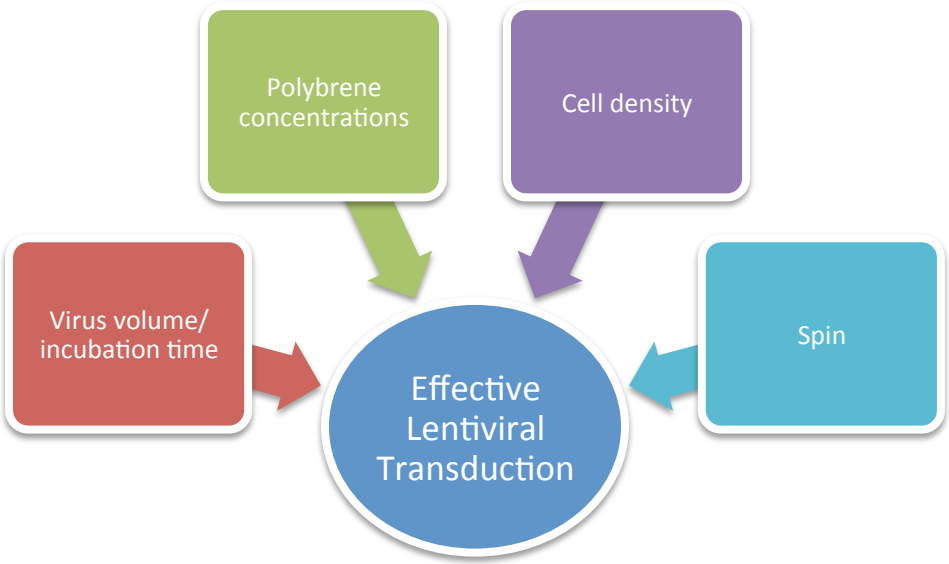
Zero virus. Minimum dose to kill? Time to kill?



Optimize library infection conditions

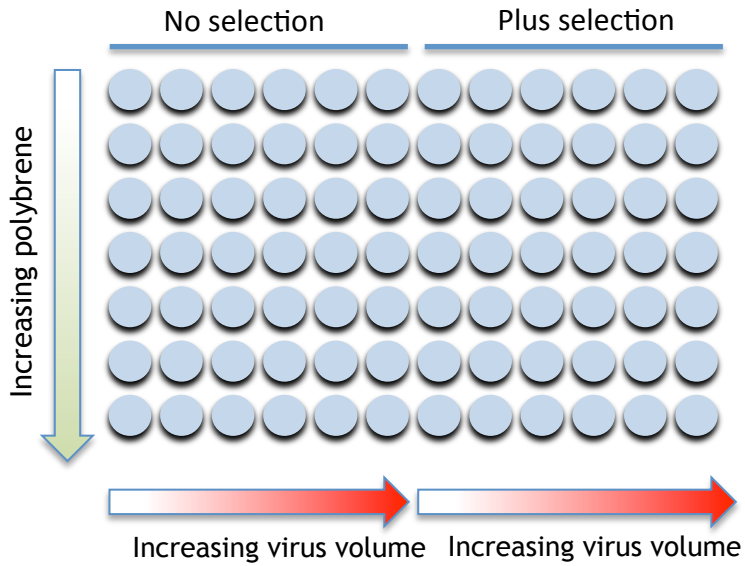


Optimize viral infection conditions –
Usually tried it already to inform cell type selection



Key variables

Matrix experiment: pre-defined

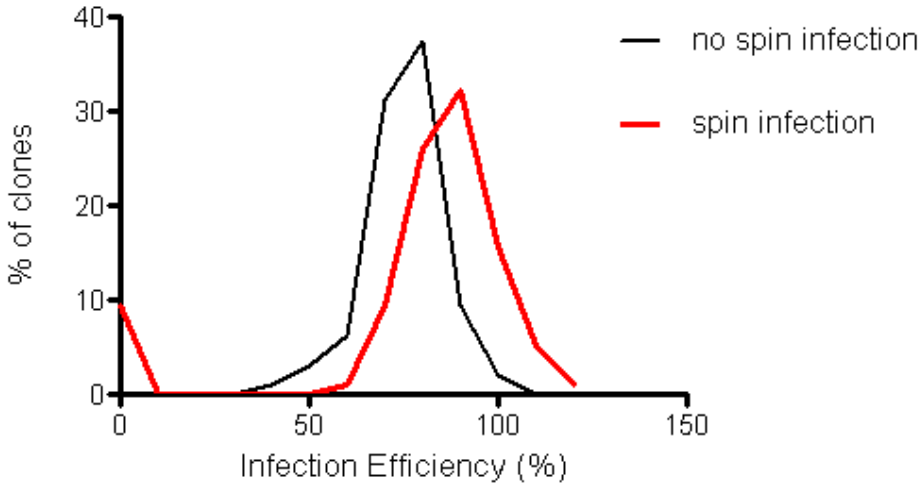


Optimize for infection rate, but also considering:

- Viral toxicity observed?
- Readout behaves comparably in treated v. untreated cells?

Optimize library infection conditions

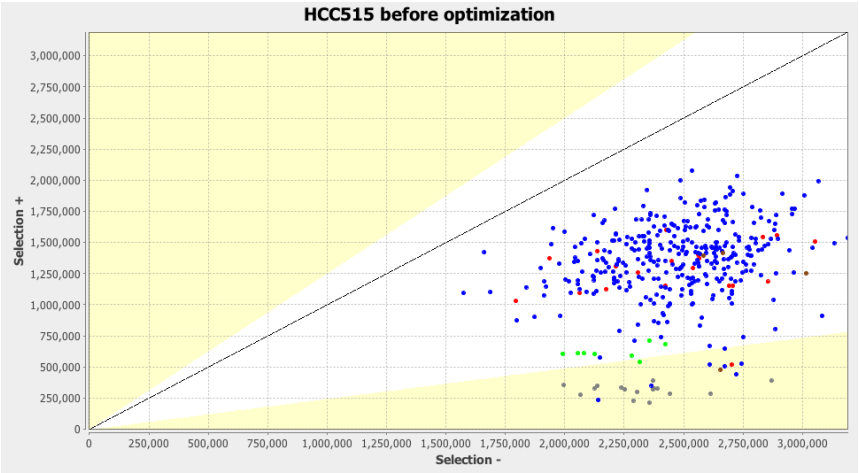
E.g. another variable: spin versus no spin

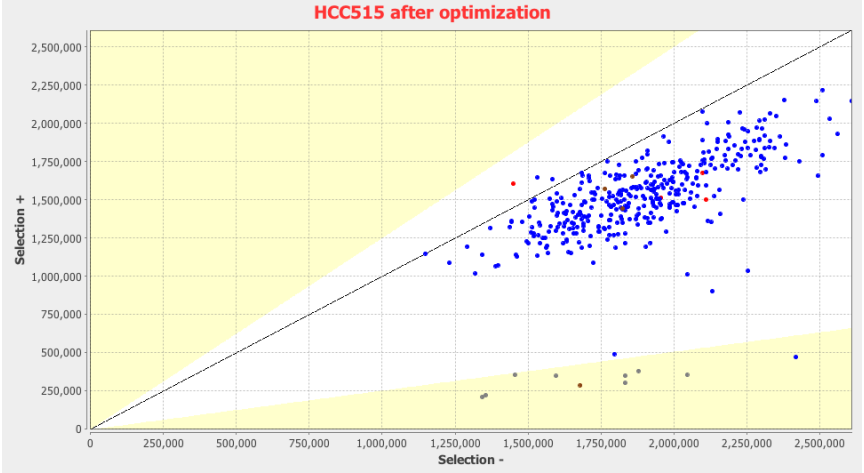


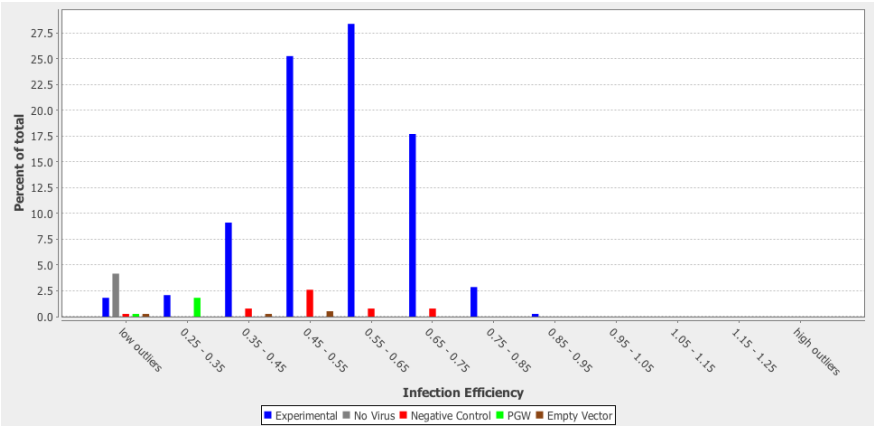
Statistical assessment – these are quantitative experiments

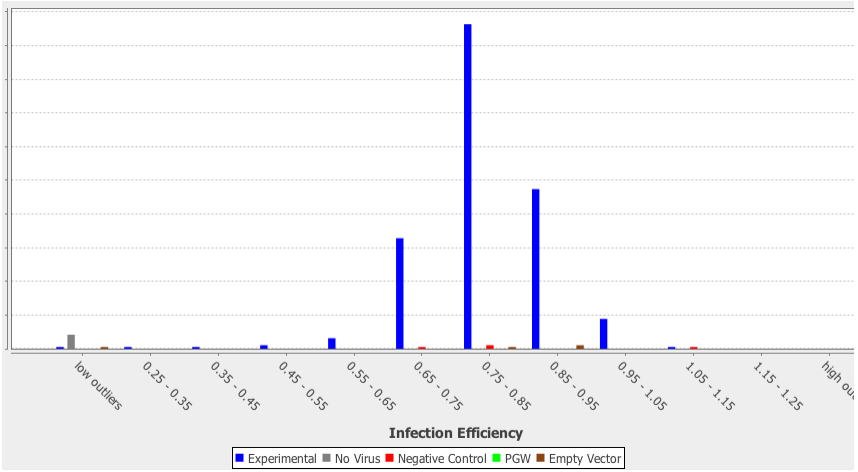
Infection optimization assessment

Before and after – compare cell viability with versus w/o selection

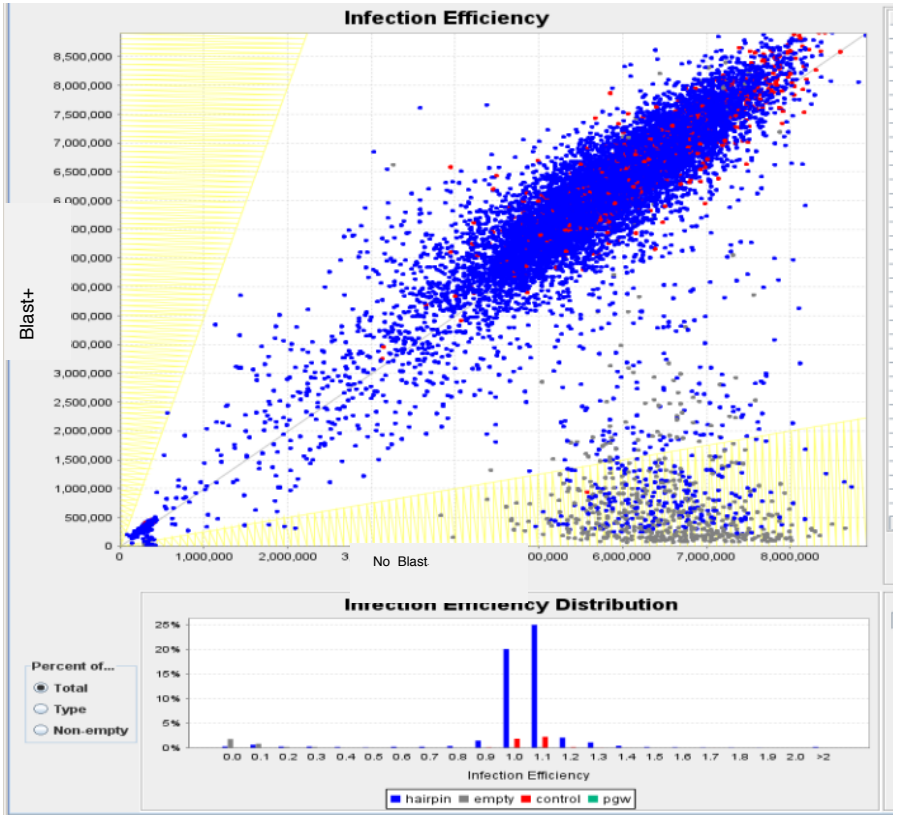




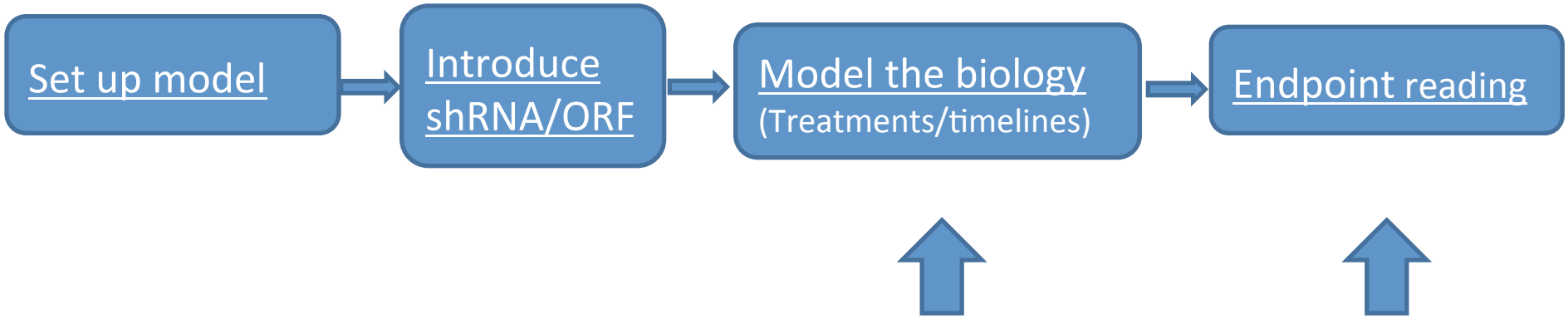




A good infection result with ORF lentiviruses



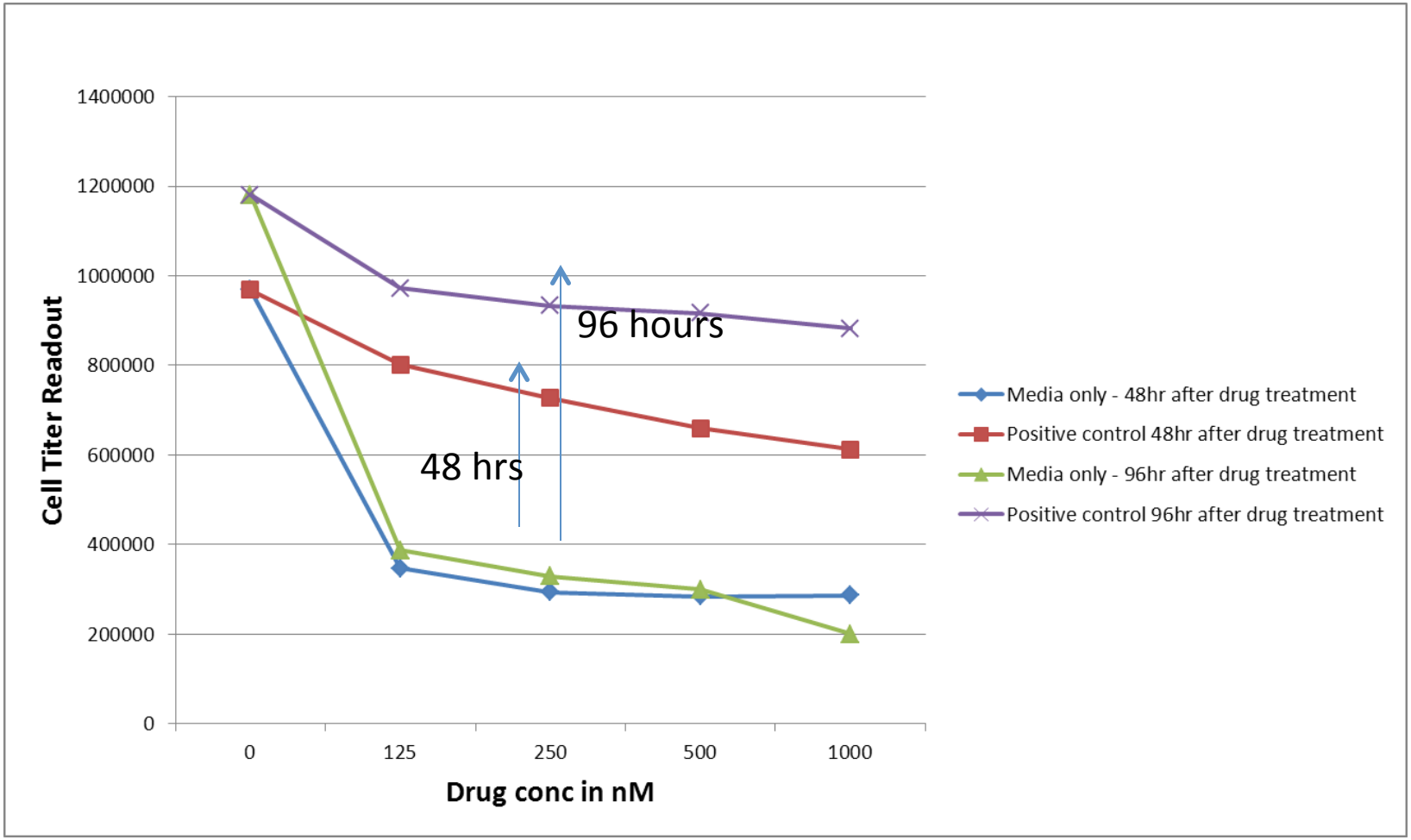
96.7% of ORFs yielding greater than 65% infection efficiency



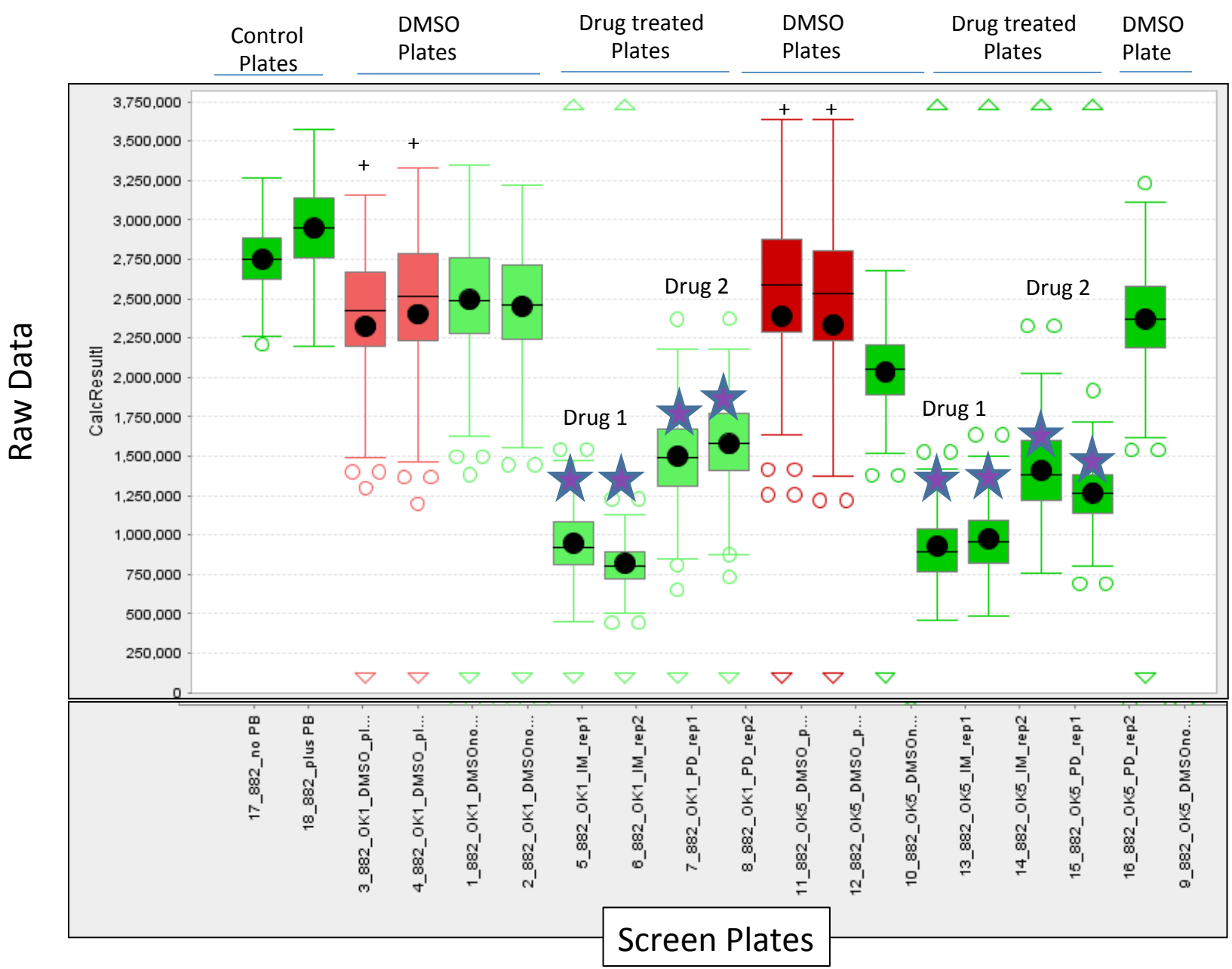
Often optimize these together

A ‘rescue’ screen: Optimizing drug dose and timepoint

This case: background = stasis by chemotherapeutic, hits = resistance



Best rescue at ~250nm at 96 hrs. Will it work? What else do you need to know?



+ DMSO Plus Selection

Positive control

Will it work?

Many assay endpoint readouts

Cell viability

Cell Titer Glo, Alamar Blue, Imaging

Reporters

Luciferase, GFP

Gene expression

qRT-PCR

L1000 Nanostring

GE-HTS (Luminex)

High Content Imaging

GFP, RFP, antibodies

Morphology, intensity, localization

Low Content "Imaging"/FACS

GFP, antibodies

Intensity

Many optimization issues are specific to readout

Always need adequate S/N

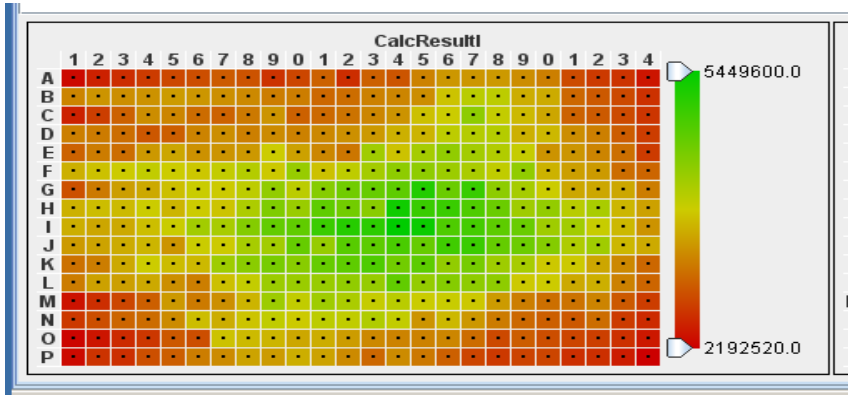
- Neg. control only ('model' positives?)
- Neg. and pos. control

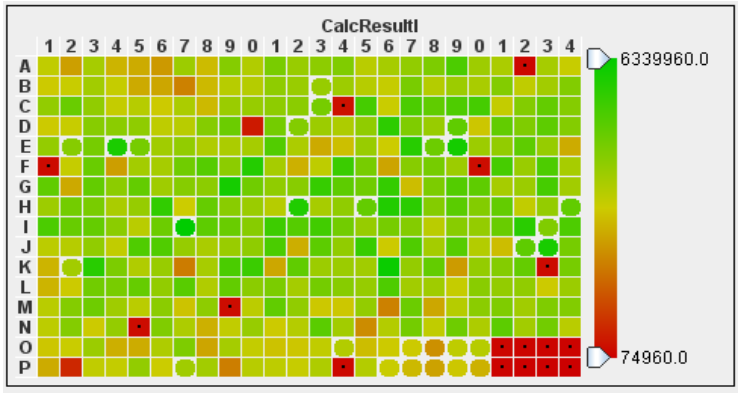
Beware artifacts – spatial effects within plates

BEFORE

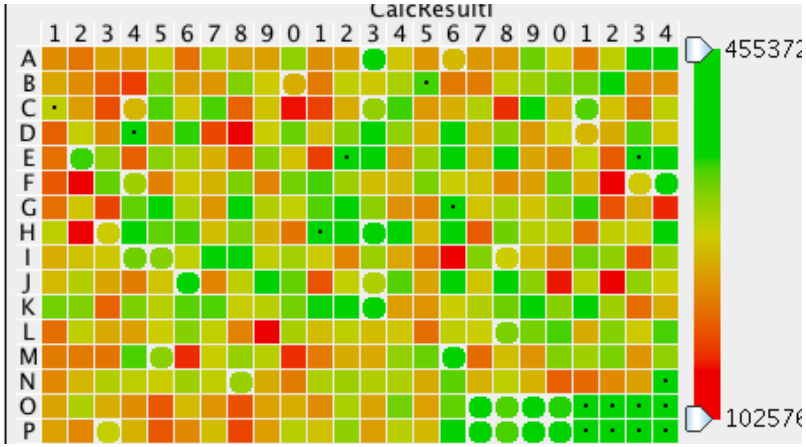
AFTER

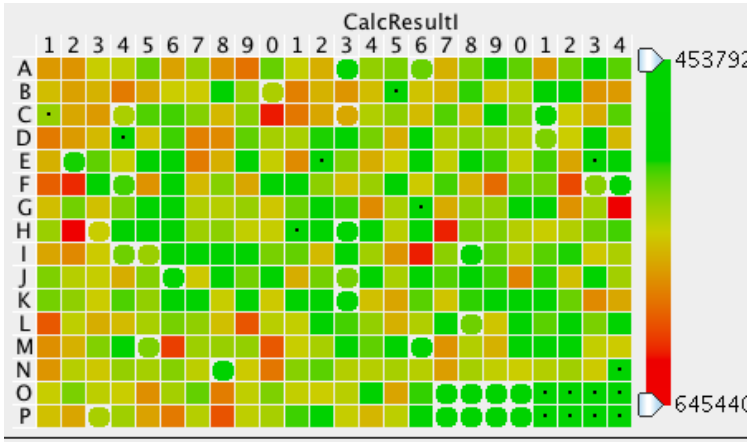
STRONG





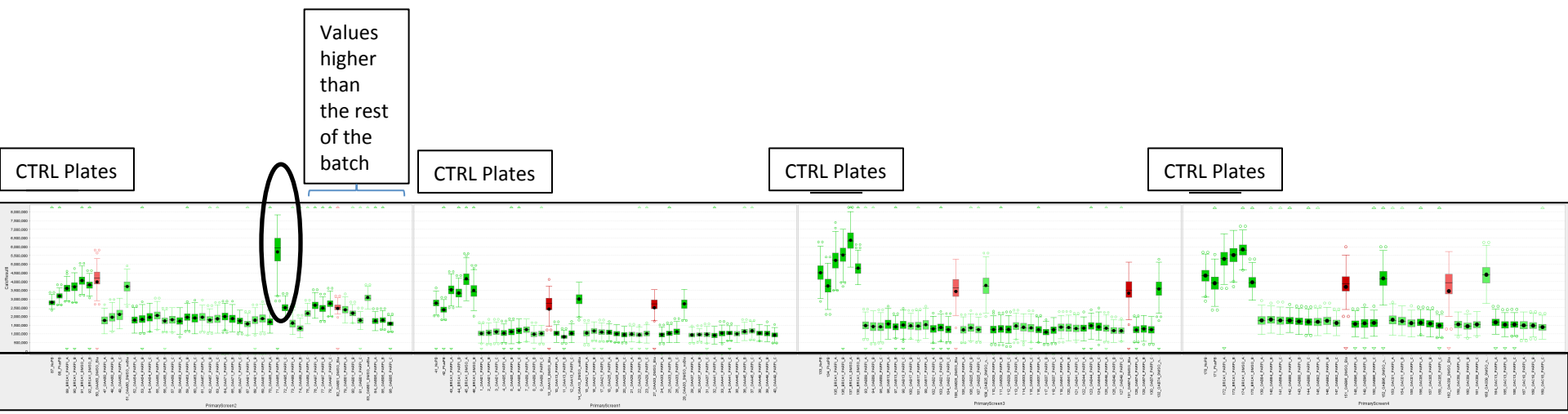
SUBTLE



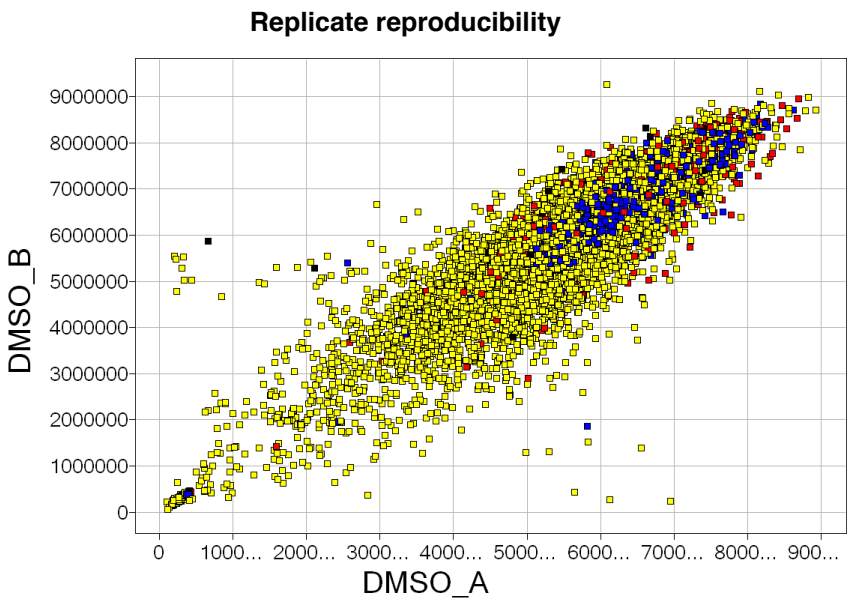


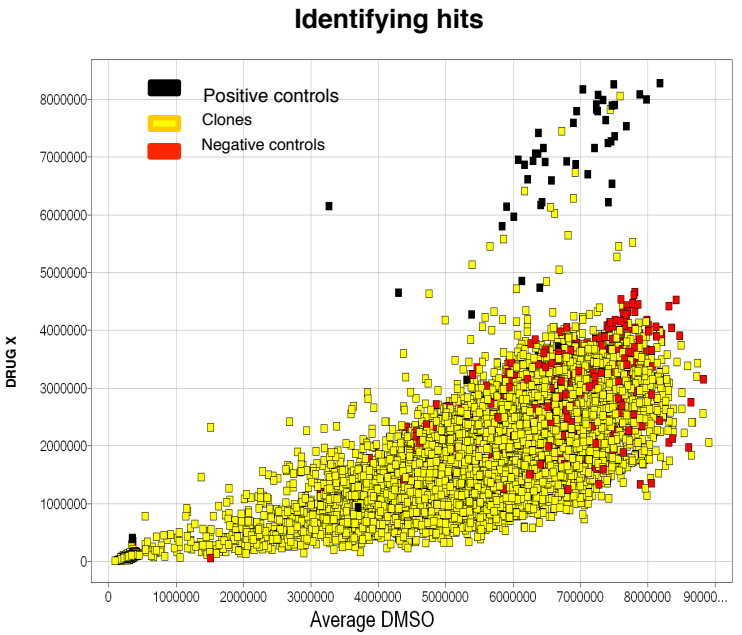
Edge effects can be reduced by optimizing the seeding, incubation, media conditions

Beware artifacts – ‘batch’ effects



Assessment of the overall optimized conditions





Replicate reproducibility:
Must be tight compared to effect size sought

Distribution of test wells, negative controls, positive controls

Pooled Screening

Multiplexed 'pooled' screens

POOLED shRNA SCREENS

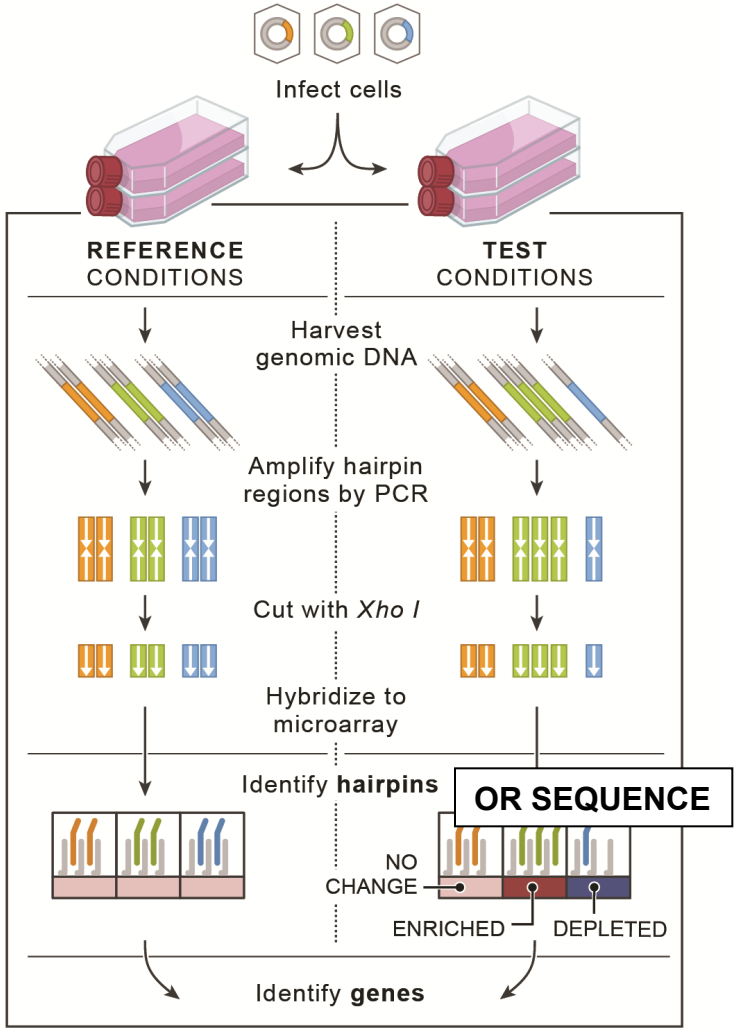
No robotics

100,000 shRNAs

- VERY LARGE GENE SETS

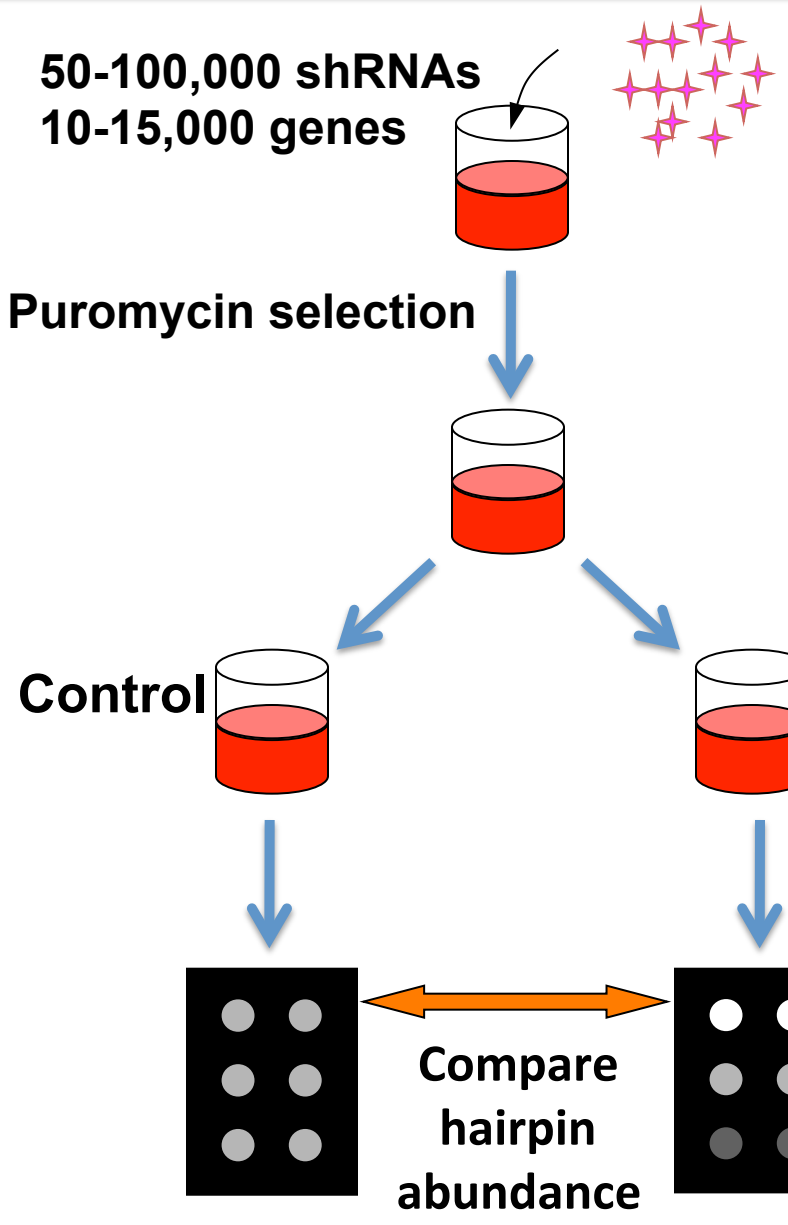
- LONG TIME COURSE ASSAYS

- MANY CONDITIONS: drug doses, time points, cell types etc.



B. Luo et. al. PNAS 105 (51) 20380 (2008)

Pooled Screening Approach



Infection procedure

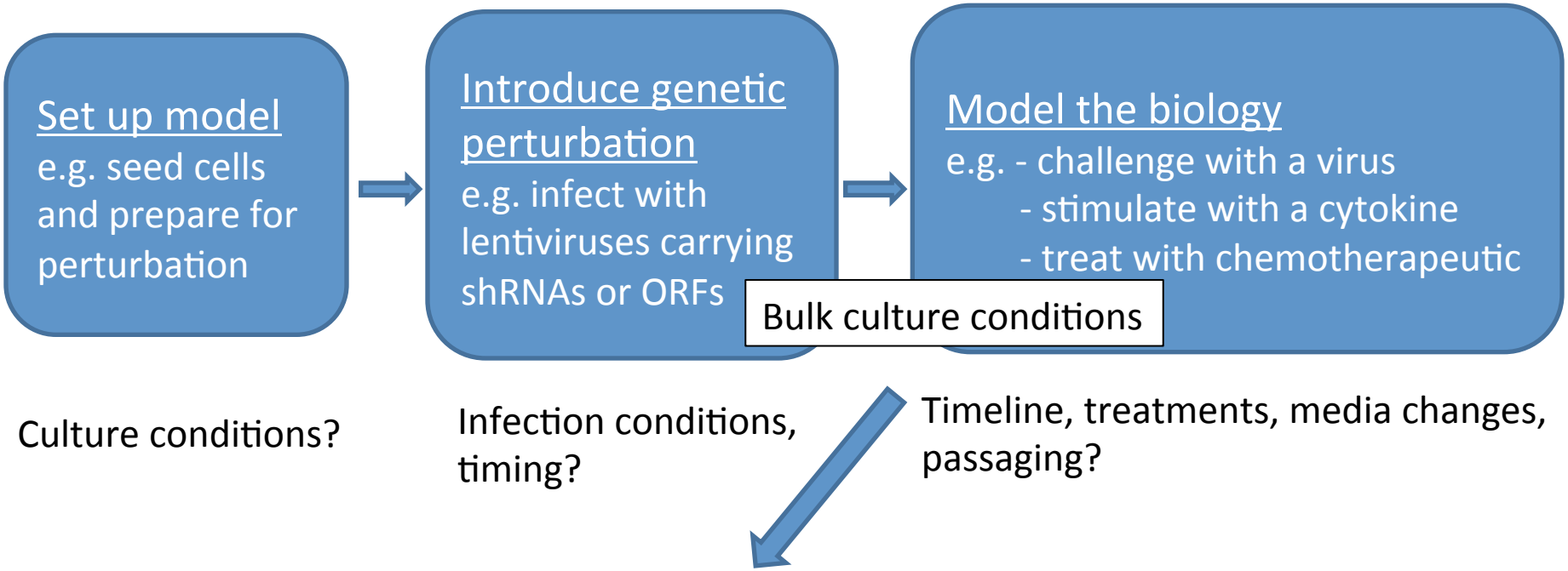
- Equal representation of each shRNA
- Infect >200 cells / shRNA
- 4 – 10 replicates

Selection procedure
maintain representation throughout experiment

→ ***Deep sequencing***

B. Luo et. al. PNAS 105 (51) 20380 (2008)

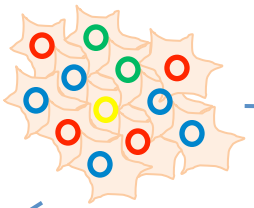
Assay Development: Many variables - many may matter!



Key: Must separate out cells with hit phenotype

- Every perturbed cell has mixed neighbors: Only a cell-autonomous effect will enrich or deplete cells carrying the shRNA/ORF that causes it.
- Need to <u>physically separate</u> 'hit' cells from others

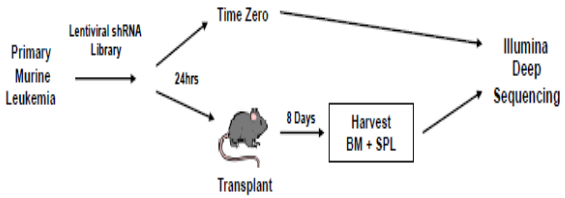
Infected cells



In vitro Proliferation

- Negative selection (Achilles)
- Chemical modifier
  - Resistance or sensitizing
- Genetic modifier

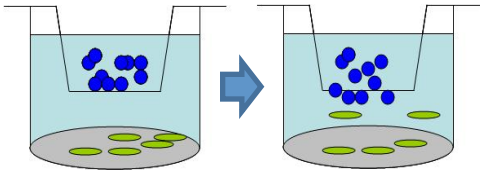
(In vivo)



Same separations as in vitro

- Proliferation
- FACS
- 'Migration' – harvest location

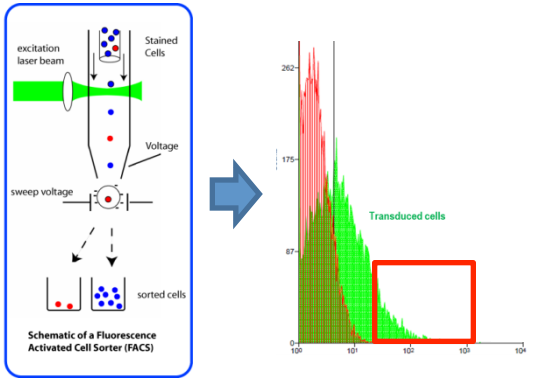
Migration



Boyden Chamber

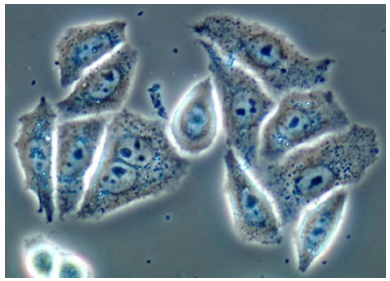
Sequence cells from both chambers to look for enrichment with phenotype of interest

FACs

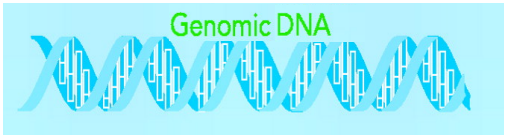


Cell surface marker
Intracellular marker (fixed)
Reporter gene

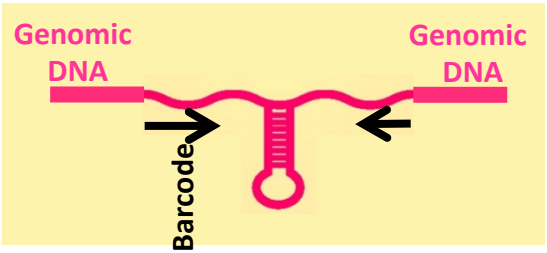
Processing of samples for Illumina deconvolution



Isolate cells enriched for phenotype of interest
→ Process enough cells to maintain representation



Isolation of gDNA



PCR of hairpin from gDNA with samples barcoded during PCR process
→ Amplify from sufficient gDNA to maintain representation

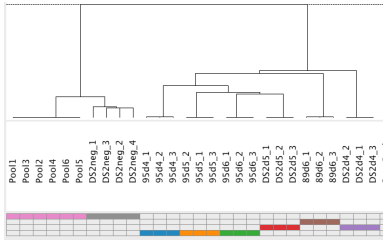
Pool all samples from same experiment into a single mix for Illumina .

Gel purify

Illumina Sequence
Run sufficient lanes to maintain representation of initial infection

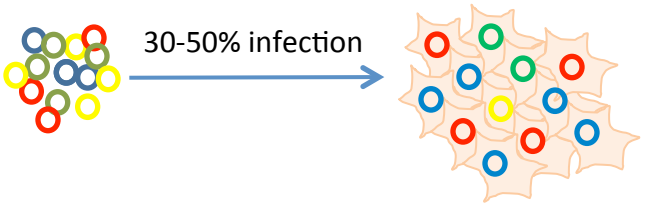
'Deconvolute' with PoolQ

Data QC



Representation considerations

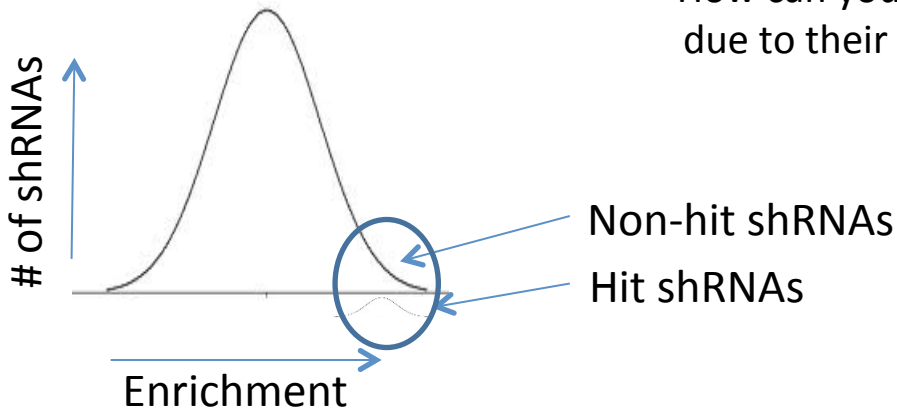
Use enough cells, or you'll get fooled



At least 200 cells per shRNA (or ORF)

1. An average – some shRNAs start out under-represented
2. Need good statistics on each shRNA

- Some will always become enriched by chance.
- How can you tell which are enriched by chance vs. due to their activity in the cells?



A "bottleneck" anywhere in the process increases random enrichments/depletions, e.g.

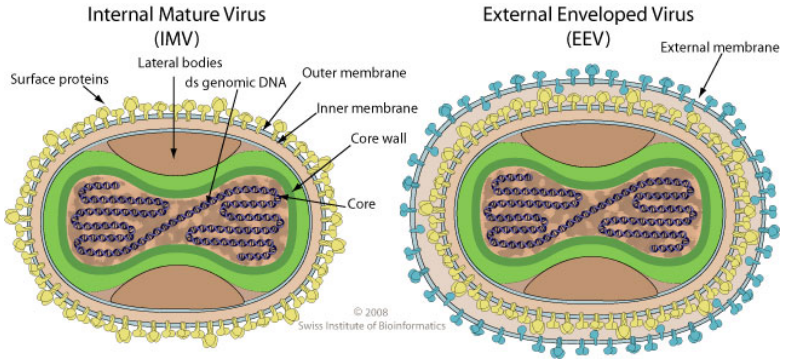
- Infect insufficient cell number
- Passage insufficient cell number

FACs based Pooled Screens
Lessons learned from case studies

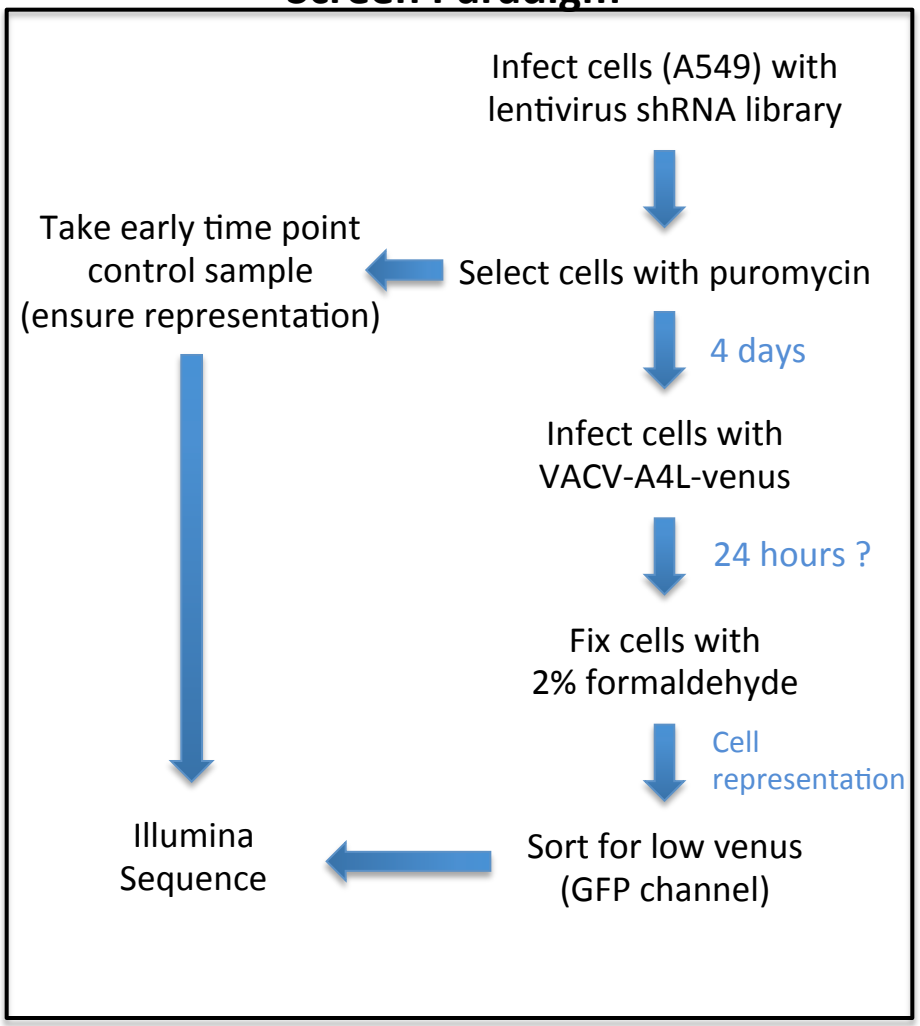
1. Orthopox viral entry/infection (Claire Marie Fillone – Connor Lab)
2. AML Differentiation (David Sykes – Scadden Lab)
3. Fetal Hb switch (Dan Bauer – Orkin Lab)

Screen 1: Orthopox viral Entry/Infection

- The orthopoxvirus family includes smallpox (Variola), Monkeypox and Vaccinia
- Smallpox was eradicated using vaccinia, but is still considered a bioterror threat
- Vaccinia is an easily manipulated model virus used under BSL-2 conditions
- Monkeypox is currently emerging in Africa
  - Imported into the United States in 2003
  - Case fatality from 1-10%

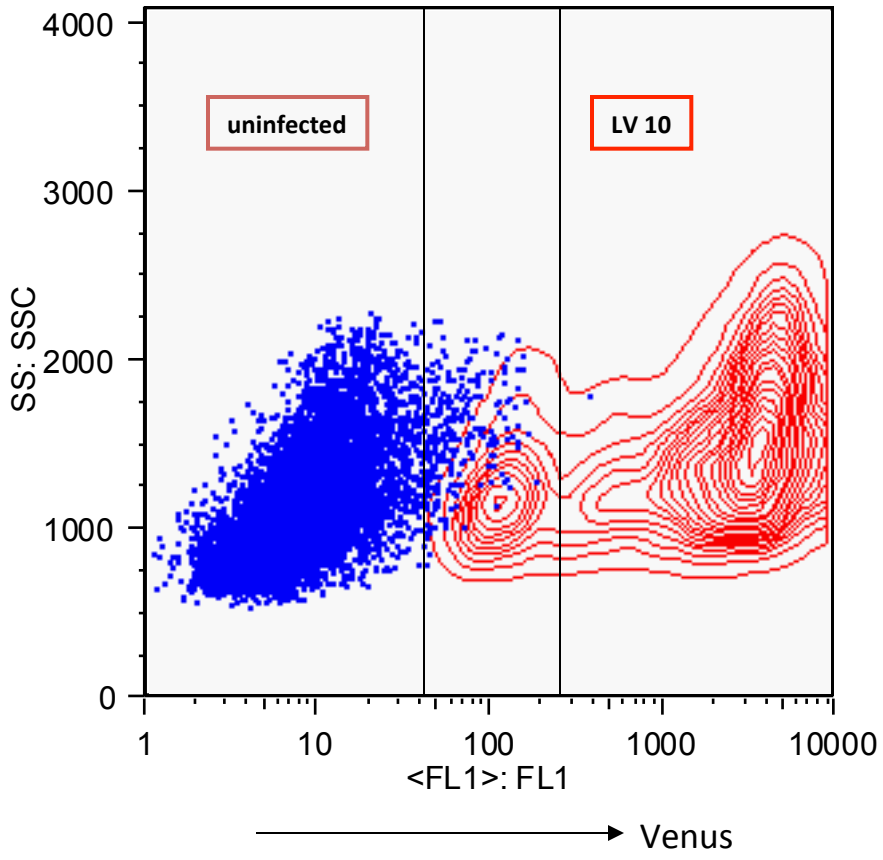


Screen Paradigm



Reporter virus testing on the BU MoFlo

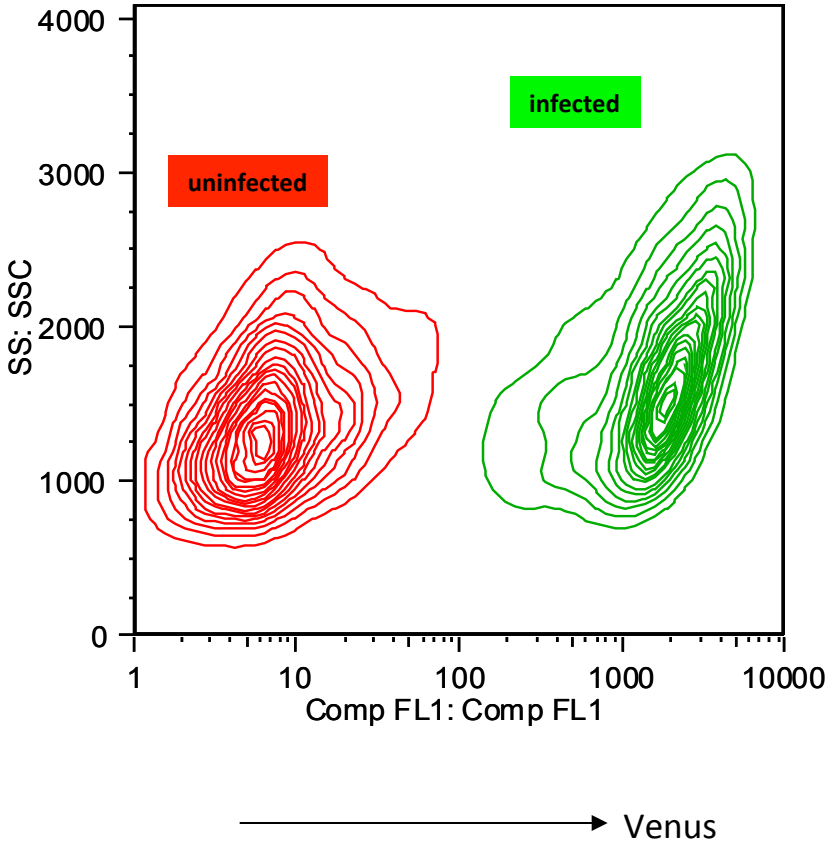
Virus expressing Venus



A549 uninfected v. LV MOI 10

Some overlap of Venus$^-$ and Venus$^+$ populations

Virus expressing viral protein **FUSED** to Venus

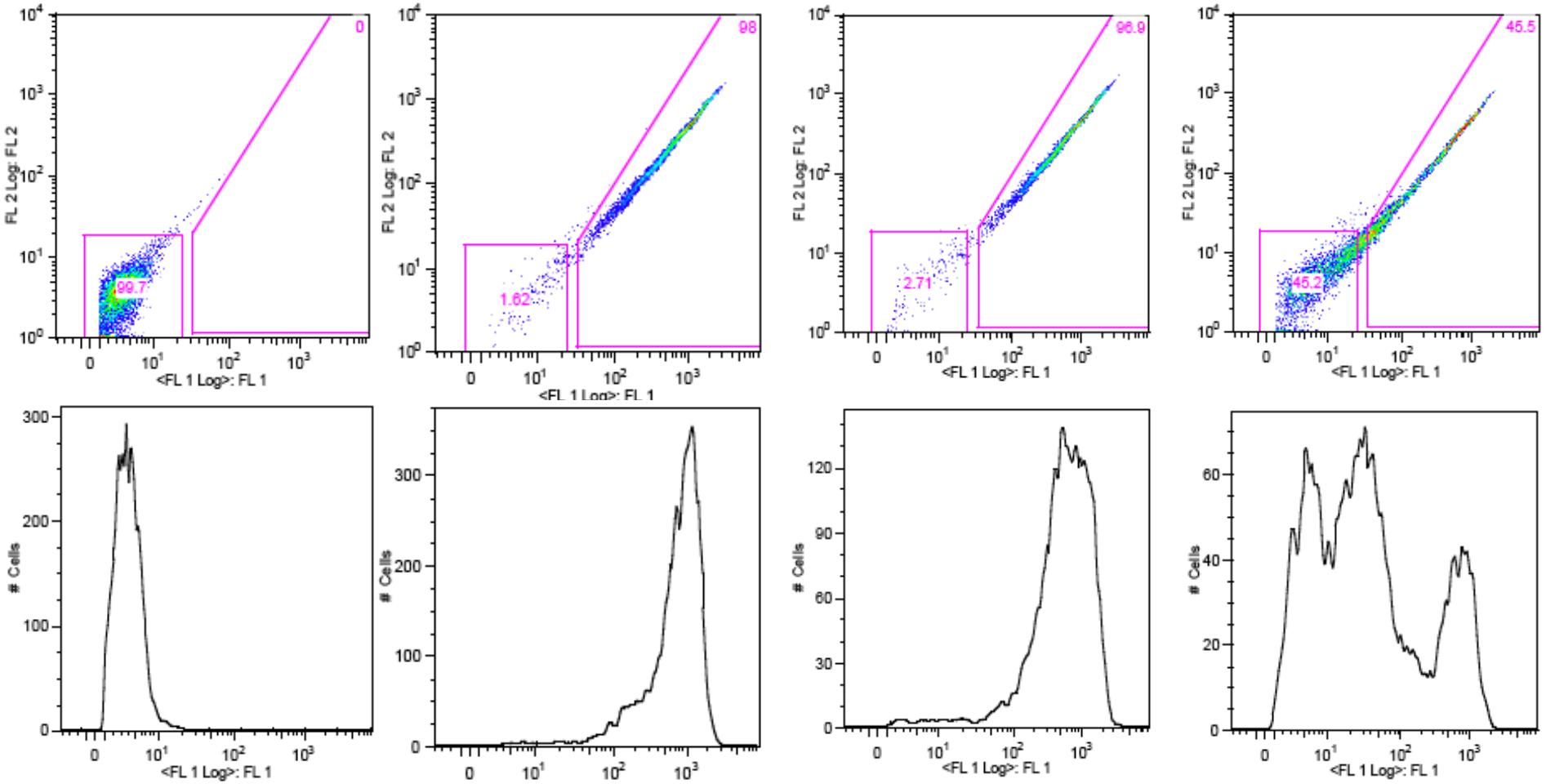


A549 uninfected v. A4L-V MOI 10

Separation of Venus$^-$ and Venus$^+$ populations

VACV A4L-Venus Test Sort

Take only 5%



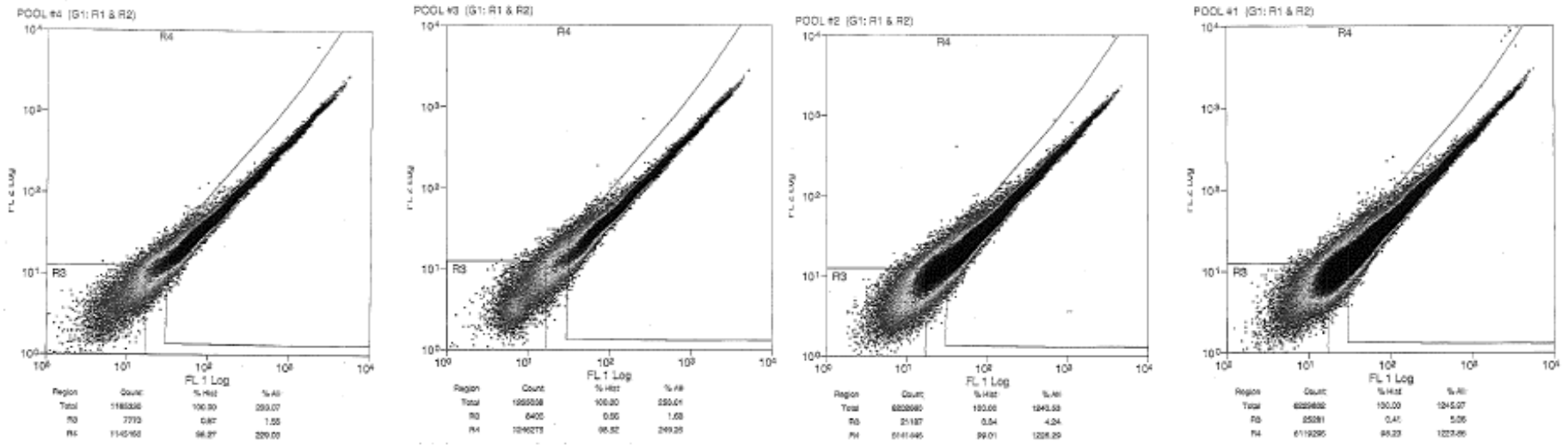
Uninfected
A549 cells

Infected w/ pLKO-empty-T
Infected w/ VACV-A4L-venus

Infected w/ pLKO-Luc
Infected w/ VACV-A4L-venus

Infected w/ pLKO-Positive shRNA at low MOI
Infected w/ VACV-A4L-venus

Comparison of 4 Replicates of Sort 90K shRNA Pool

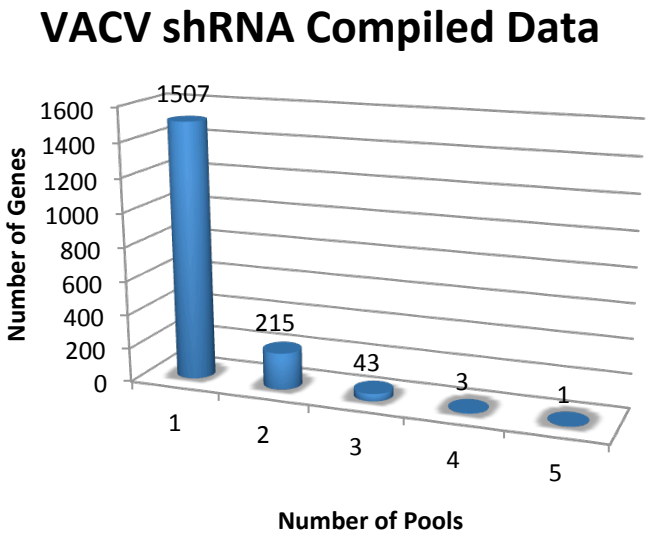


- Pool #1: 3.71×10^7 total, 23,209 sorted
- Pool #2: 3.48×10^7 total, 18,457 sorted
- Pool #3: 1.02×10^8 total, 220,694 sorted
- Pool #4: 5.68×10^7 total, 28,096 sorted

Capturing this info can help in flagging replicates in downstream analysis

Primary screen replicate analysis
Choosing genes for follow-up

- Identified shRNA rank list for each pool, and a union of all 4 pools (5 'pools')
- Used GENE-E 2nd Best Rank to nominate genes from ranked hairpin list
- Comparison of number of genes found in multiple pools



- Chose 170 genes (10 re-array plates)
  - All genes in 3 or 4 pools
  - 20 genes from top of each individual pool
  - 5 genes nominated by having top hairpin in each pool
  - 25 genes in 2 pools, excluding the sum 'pool'

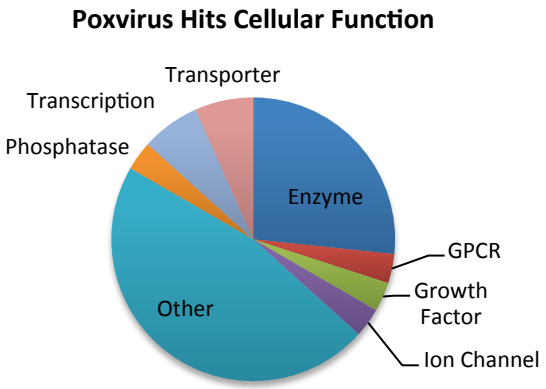
For a very stringent selection screen, allow for a big false negative rate, i.e. don't ask for as much consistency across trials

Follow up confirmation rate (Array-based Screen)

Very stringent criteria for 'confirmation'.

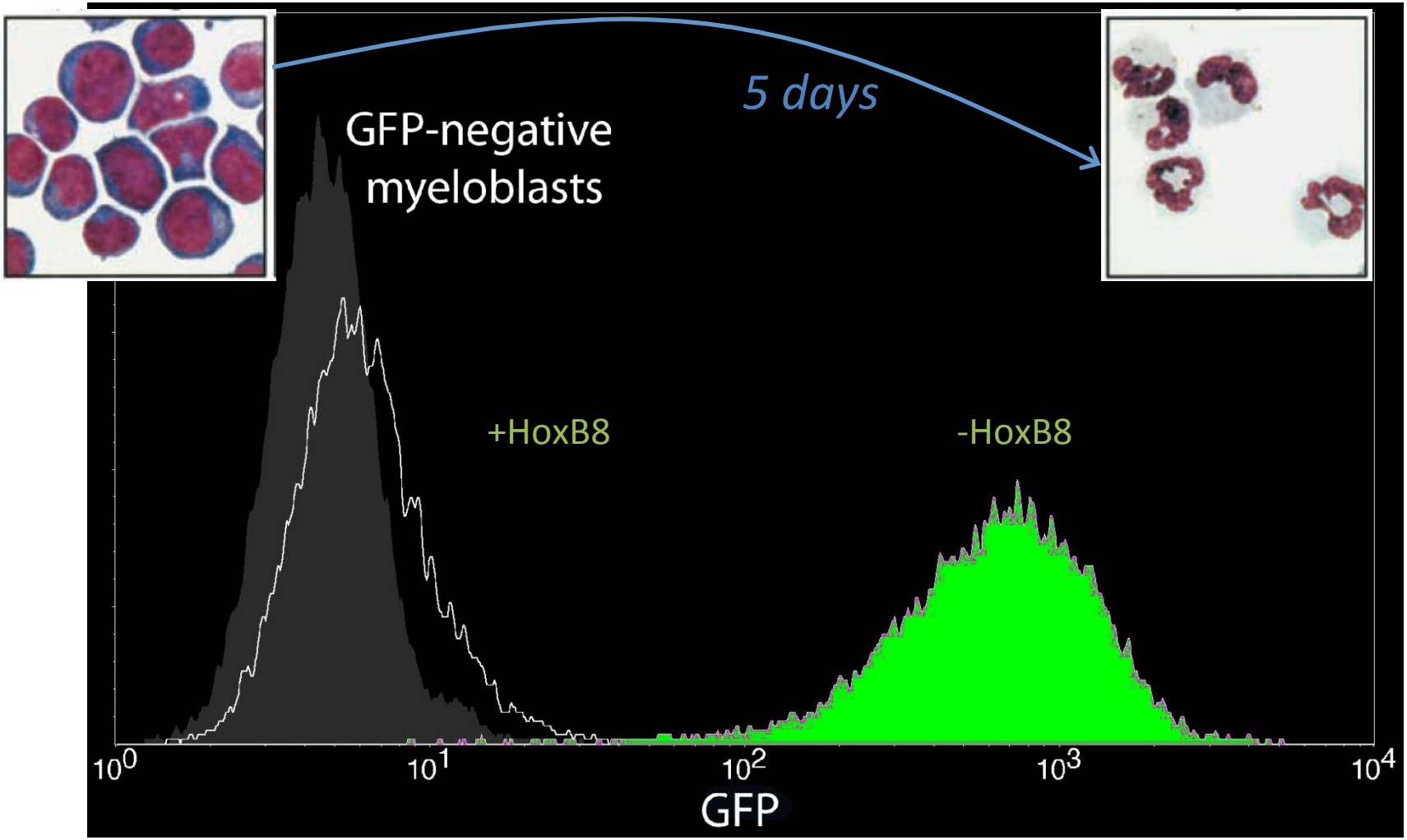
- To be considered inhibited, need to block 50% fluorescence expression
- To be considered a hit, the genes must inhibit in more than one replicate
  - Have 2 or more hairpins per gene

Hits: 32 of 170 genes fit this description (~20%)



What's this hit rate mean?
Not much in itself:
Don't get hung up on % hit values that use <u>arbitrary thresholds.</u>

Screen 2: AML model: HoxB8-induced differentiation block

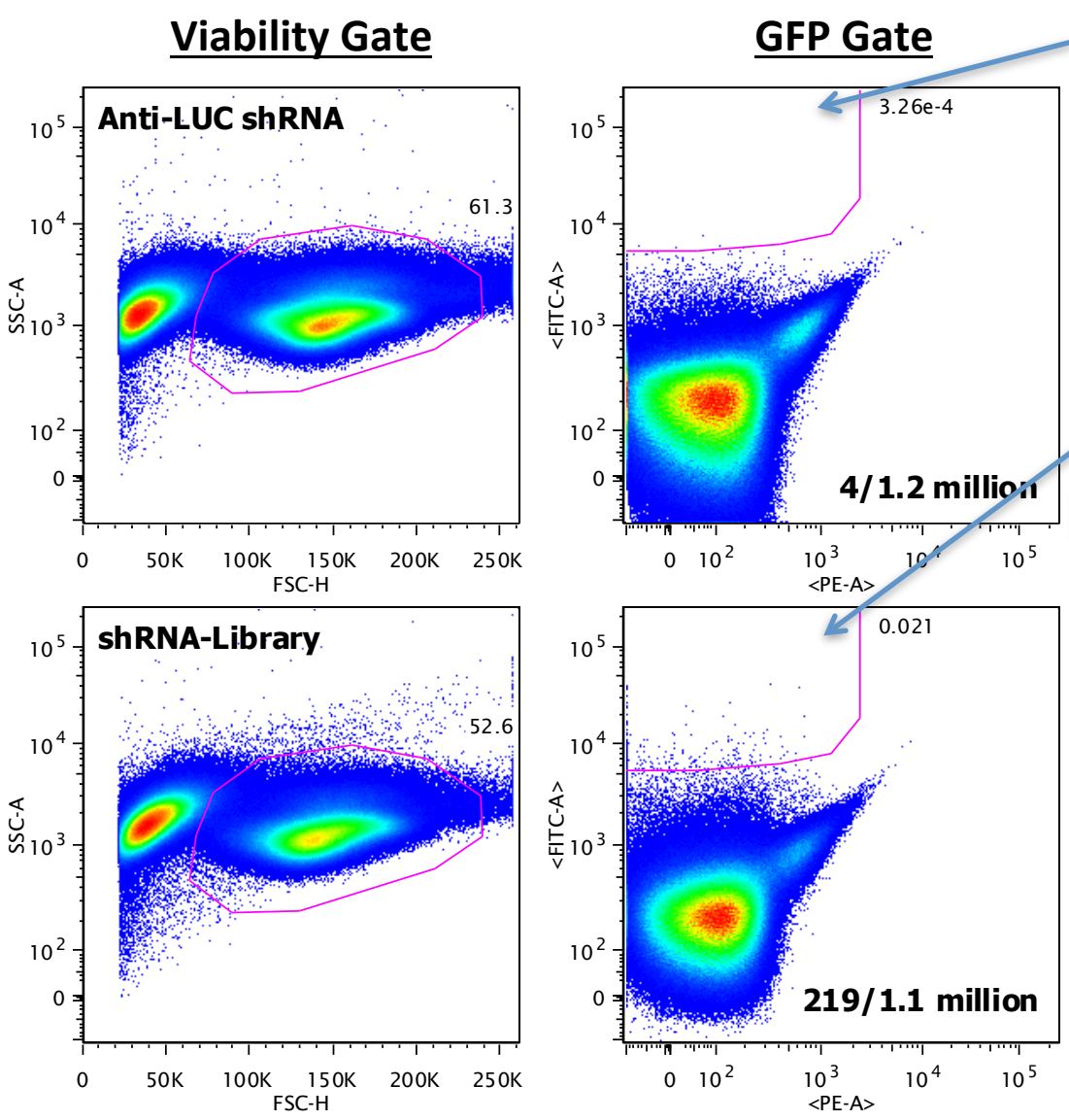


David Sykes, Scadden Lab

Flow cytometry selection of hit cells

Background: Introduce a single neg. control shRNA

Screen: Introduce 40K shRNAs



4 background apparent 'hit' cells

In actual screen, sort ~50X more cells to obtain ~10K hit cells per screen replicate

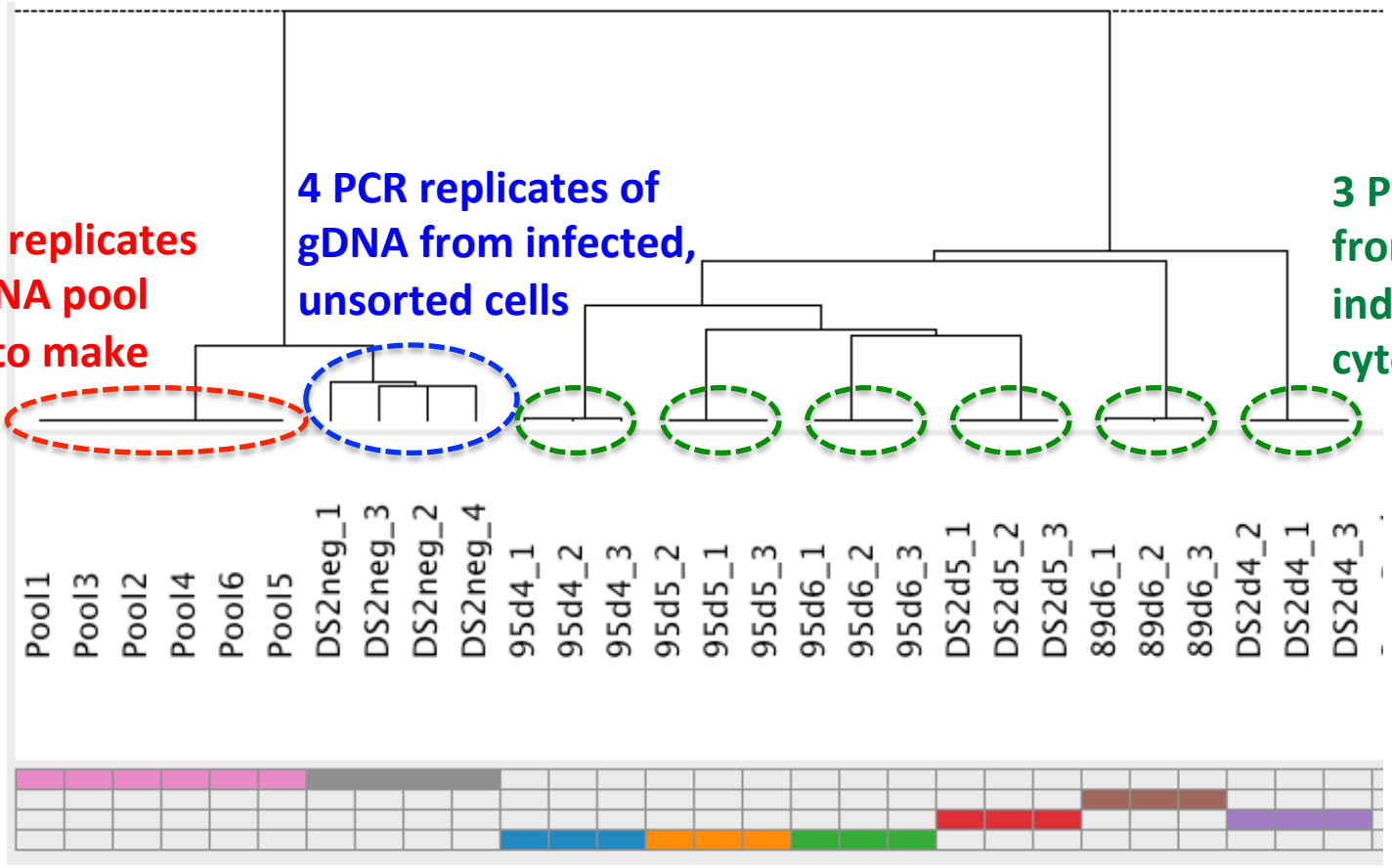
219 'hit' cells

60-fold enrichment of GFP$^+$ cells from library infection

Conditions

- Optimize for time point for flow following infection: Collect day 4, 5, or 6
  - Replicate 1: Day 4, 5
  - Replicate 2: Day 6
  - Replicate 3: Day 4, 5, 6
- 200 shRNAs per cell

Clustering of indexed Illumina reads

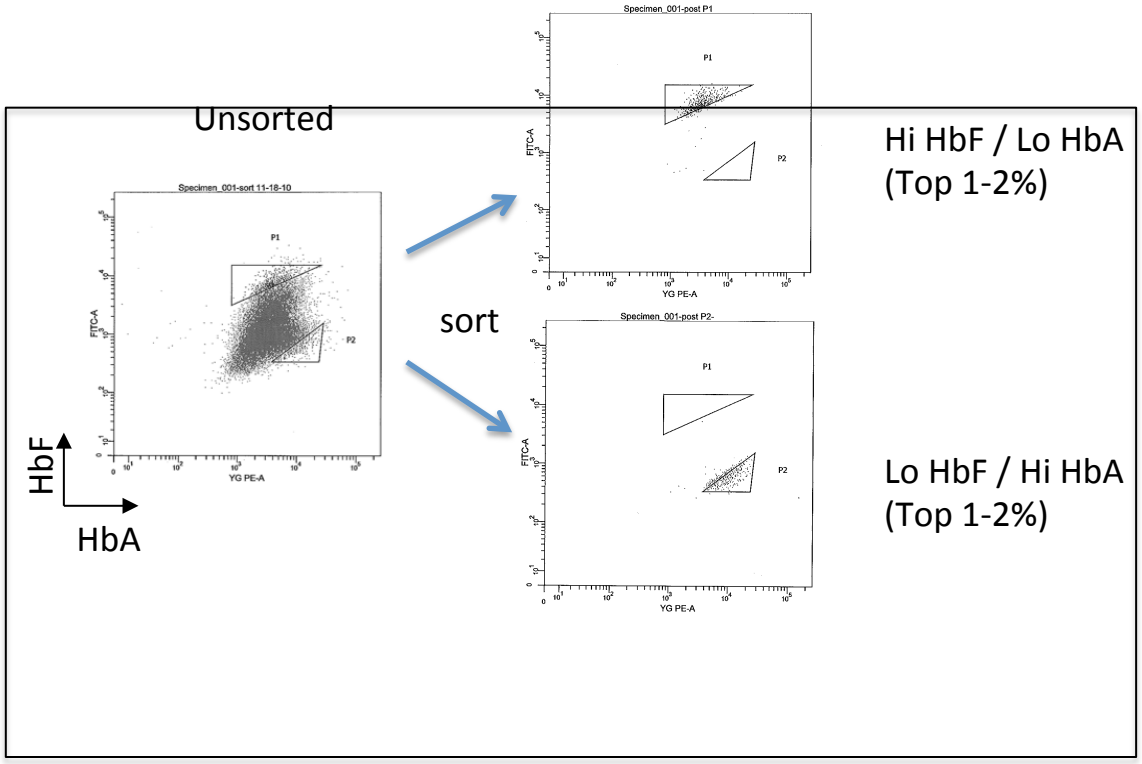


Top hit makes sense, strong miRNA signal too: screen worked!

Rank	Symbol	Sequence	Counts per million		Fold Enrichment
			Unsorted	GFP+	
1	Hoxb8	UAGCCGUAGAAGUUGCCGUUUU	0.1	2185.1	**36490**
2	Traf5	AAU**UCUCUC**AGAGACCGGUUUU	1.1	2118.8	**1981**
3	LOC434093	GUGUUGACUAUACAGCCGUUUU	1.0	476.7	**482**
4	1810035L1	GU**UCUCUC**AGCUCACUCGUUUU	1.2	548.2	**444**
5	LOC381842	G**UCUCUC**UUACUGGUAGGUUUU	110.8	22186.2	**200**
6	Itgax	U**UCUCUC**UGCAUGUGUGGUUUU	39.3	7362.4	**188**
7	Ehbp1	**AUUUGG**CUUUGUGAUAGCUUUU	36.3	6245.2	**172**
8	Eraf	**AUUUGG**CUAGAAACUGGCUUUU	39.6	6778.0	**171**
9	Oprd1	A**AUUUGG**UGUACCGGACGUUUU	8.2	1323.5	**161**
10	Slc2a8	AU**UCUCUC**UUCUACCUGGUUUU	11.4	1804.0	**159**
102	Hoxb8	ACUGCUGGGAAACUUGUCUUUU	22.6	593.7	**26**

3. A two-color FACS screen:

Sort on both Adult and Fetal hemoglobin level.
Screen for high OR low ratios.

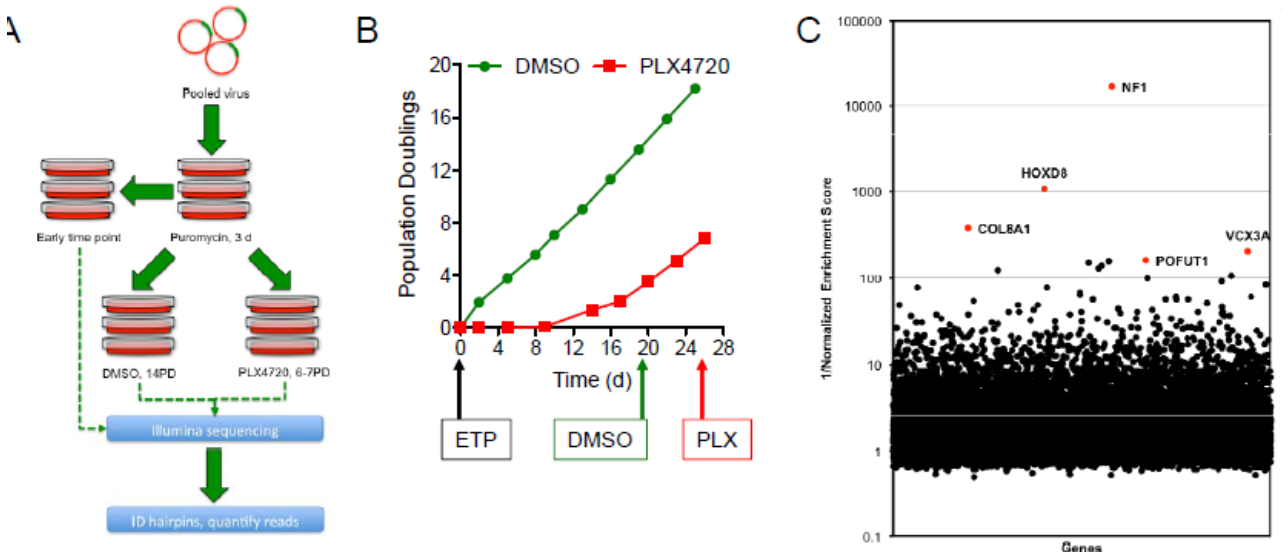


Bauer, Orkin

NF1 loss as a mechanism of resistance to selective RAF and MEK inhibition

1. Melanoma cell lines with BRAF V600E mutation are sensitive to PLX 4720

2. Perform genome-wide shRNA pooled screen in the presence of a BRAF inhibitor to identify genes that cause resistance to cells treated with drug



Modifier screens need to determine cell line choice, correct dosage, and time-points as main variable

Steven Whittaker

Some pooled screen take-home lessons

1. Need very low background of 'hit' cells
2. Avoid any BOTTLENECK in cell numbers in any pooled screen.
3. Even w/o positive control, can look for existence of a hit population by increase in hit cells w library treatment vs. control shRNA.
4. Replicates do not have to cluster if screen is sub-saturating? (I.e. not enough representation of hits) Need larger number of replicates (ie just more cells)
5. Primary cell screens are possible, but may be more difficult
6. For FACS-based screens, if using a florescent reporter and fixing cells, fuse report to endogenous protein to avoid leakage
7. For FACS-based screens, ensure sort maintains representation (avoid long sorts where cells can clump) (see #2)
8. For FACS-based screens, re-sorting (Sort-expand-Sort) did not increase signal over background (one test)

Assay Development: Many variables - many may matter!